

**ADVANCED PROTEOMIC CHARACTERIZATION OF THE 26S PROTEASOME IN  
ARABIDOPSIS REVEALS INSIGHTS INTO COMPOSITION AND ASSEMBLY**

by

David C. Gemperline

A dissertation submitted in partial fulfillment of  
the requirements for the degree of

Doctor of Philosophy

(Genetics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: TBD

The dissertation is approved by the following members of the Final Oral Committee:

Richard D. Vierstra, Professor, Genetics  
Richard Amasino, Professor, Genetics  
Josh Coon, Professor, Chemistry  
Donna Fernandez, Professor, Botany  
Patrick Masson, Professor, Genetics

© Copyright by David C. Gemperline 2016

All Rights Reserved

*To Erin, my wife.*

## ACKNOWLEDGMENTS

*Science doesn't purvey absolute truth. Science is a mechanism, a way of trying to improve your knowledge of nature. It's a system for testing your thoughts against the universe and seeing whether they match. This works not just for the ordinary aspects of science, but for all of life.*

— ISAAC ASIMOV (1988)

Acknowledgements go here.

**TABLE OF CONTENTS**

|   |      |
|---|------|
| Table of Contents   | iii  |
| List of Tables  | v    |
| List of Figures   | vi   |
| List of Abbreviations and Acronyms                                | viii |
| Abstract  | ix   |
| Abstract  | xii  |
| Chapter 1: The Ubiquitin 26S Proteasome System                    | 1    |
| 1.1 Introduction to the Ubiquitin 26S Proteasome System . . . . . | 1    |
| 1.2 The 26S Proteasome . . . . .                                  | 2    |
| 1.3 26S Proteasome Assembly . . . . .                             | 17   |
| 1.4 Proteasome Post-Translational Modification . . . . .          | 17   |
| 1.5 Proteasome Degredation . . . . .                              | 17   |
| 1.6 Proteasome Interacting Proteins . . . . .                     | 17   |
| 1.7 References . . . . .  | 17   |

|   |    |
|---|----|
| Chapter 2: Morpheus Spectral Counter: A Computational Tool for Label-Free Quantitative Mass Spectrometry using the Morpheus Search Engine | 36 |
| 2.1 Summary . . . . .   | 36 |
| 2.2 Main Text . . . . .   | 37 |
| 2.3 Methods . . . . .   | 51 |
| 2.3.1 Sample Preparation . . . . .  | 51 |
| 2.3.2 Liquid Chromatography and High-Resolution Mass Spectrometry . . . . .   | 53 |
| 2.3.3 Data Processing . . . . .   | 54 |
| 2.3.4 Isoform Incorporation Rates . . . . .   | 56 |
| 2.3.5 Speed and Accuracy Comparisons of Morpheus/MSpC to the TPP/ABACUS Pipeline . . . . .  | 56 |
| 2.4 Discussion on Requiring Two Unique Peptides to Quantify a Protein   | 58 |
| 2.5 Tutorial . . . . .  | 60 |
| 2.6 Acknowledgements . . . . .  | 64 |
| 2.7 References . . . . .  | 64 |
| Colophon  | 69 |

**LIST OF TABLES**

|  |    |
|--|----|
| 2.1 Table of dNSAF values for each 20S proteasome subunit generated by<br>analyzing the proteasome spike in experiments with the Morpheus and<br>MSpC pipeline . . . . .                       | 49 |
| 2.2 Table of adj_NSAF values for each 20S proteasome subunit (equivalent<br>to dNSAF) generated by analzying the proteasome spike in experiments<br>with the TPP and ABACUS pipeline . . . . . | 50 |

## LIST OF FIGURES

|     |  |    |
|-----|--|----|
| 1.1 | The Ubiquitin 26S Proteasome System . . . . .  | 3  |
| 1.2 | Structure of the 26S proteasome . . . . .  | 4  |
| 1.3 | Electron microscopy images of the 20S and 26S proteasomes from mammals and plants . . . . .  | 10 |
| 1.4 | Affinity purification of 26S proteasomes from <i>pag1-1 PAG1-FLAG</i> plants   | 16 |
| 2.1 | MSpC Graphic User Interface (GUI) and software flow chart . . . . .  | 39 |
| 2.2 | Confirmation of MSpC accuracy by analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2) . . . . .  | 41 |
| 2.3 | Re- Analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2) . . . . .   | 43 |
| 2.4 | Confirmation of MSpC accuracy by analysis of MS/MS datasets generated with affinity purified <i>Arabidopsis</i> 20S proteasomes spiked into a total cell lysate from <i>E. coli</i> . Following MS/MS analysis, the dNSAF and uNSAF values for each subunit/isoform were determined by Morpheus and MSpC . . . . . | 46 |

|  |    |
|--|----|
| 2.5 MSpC combined with Morpheus works faster than TPP combined with ABACUS. Speed comparisons were performed for 12, 24, 48, and 96 raw MS/MS files generated with the 20S proteasome/ <i>E. coli</i> lysate samples analyzed in 2.4 . . . . . | 52 |
|--|----|

**LIST OF ABBREVIATIONS AND ACRONYMS**

|          |  |
|----------|--|
| API      | Application programming interface      |
| BCA      | Bicinchoninic acid protein assay       |
| BLAST    | Basic local alignment search tool      |
| C#       | C sharp, a programming language        |
| Da       | Dalton, the atomic mass unit           |
| DNA      | Deoxyribonucleic acid                  |
| DTT      | Dithiothreitol                         |
| ESI      | Electrospray ionization                |
| E-value  | Expectation value                      |
| FASTA    | A format for storing protein sequences |
| FDR      | False discovery rate                   |
| GUI      | Graphical user interface               |
| HCD      | Higher-energy collisional dissociation |
| HPLC     | High-performance liquid chromatography |
| LC       | Liquid chromatography                  |
| <i>m</i> | Mass                                   |
| min      | Minute                                 |
| MS       | Mass spectrometry                      |

|                 |  |
|-----------------|--|
| MS <sup>1</sup> | Survey mass analysis                                   |
| MS/MS           | Tandem mass spectrometry                               |
| NCE             | Normalized collision energy                            |
| nLC             | Nanoflow liquid chromatography                         |
| ppm             | Part per million                                       |
| PSM             | Peptide-spectrum match                                 |
| PTM             | Post-translational modification                        |
| s               | Second   |
| SILAC           | Stable isotope labeling by amino acids in cell culture |
| S/N             | Signal-to-noise ratio                                  |
| TMT             | Tandem mass tag  |

**ADVANCED PROTEOMIC CHARACTERIZATION OF THE 26S  
PROTEASOME IN ARABIDOPSIS REVEALS INSIGHTS INTO  
COMPOSITION AND ASSEMBLY**

David C. Gemperline

Under the supervision of Professor Richard D. Vierstra

At the University of Wisconsin-Madison

**Abstract**

The 26S proteasome is the central proteolytic effector in the ubiquitin system that is responsible for degrading numerous regulators following their selective ubiquitylation. While much is known about the construction of the yeast and mammal particles, little is known about the pathways used to assemble the plant particle. One challenge is that the known yeast chaperones appear sufficiently diverged to preclude high-confidence identification of their plant counterparts by genomic searches. Here, we used in-depth mass spectrometric analysis of Arabidopsis 26S proteasomes, which were affinity purified from seedlings under conditions that promote the accumulation of assembly intermediates, to identify a large collection of interacting proteins that associate with either the core protease (CP) or regulatory particle (RP). Sequence comparisons, Y2H and BiFC studies revealed

that some are likely assembly chaperones, with several CP factors harboring the signature C-terminal HbYX motif that allows their association with the  $\alpha$ -subunit ring. Several of the RP-specific factors appear to be orthologs of the chaperones Nas2, Nas6, Hsm3 and Ecm29. Whereas yeast assembles only a single particle type, mammals can assemble alternate proteasomes by replacing individual subunits with distinct isoforms (e.g., immunoproteasomes). In plants, most 26S proteasome subunits are encoded by paralogous genes with sufficient divergence to suggest that plants also accumulate a collection of particles. However, proteomic analysis of proteasomes selectively enriched using paralog-specific tags strongly imply that although plants possess this genetic diversity, the incorporation of these paralogs appears random, and is mainly influenced by the differential expression of the corresponding genes. Taken together, these proteomic studies provide the first insights into plant proteasome assembly and diversity, and identify factors that build the CP and RP subcomplexes and finally the 26S holo-particle.

Richard D. Vierstra

## ABSTRACT

### Abstract

The 26S proteasome is the central proteolytic effector in the ubiquitin system that is responsible for degrading numerous regulators following their selective ubiquitylation. While much is known about the construction of the yeast and mammal particles, little is known about the pathways used to assemble the plant particle. One challenge is that the known yeast chaperones appear sufficiently diverged to preclude high-confidence identification of their plant counterparts by genomic searches. Here, we used in-depth mass spectrometric analysis of *Arabidopsis* 26S proteasomes, which were affinity purified from seedlings under conditions that promote the accumulation of assembly intermediates, to identify a large collection of interacting proteins that associate with either the core protease (CP) or regulatory particle (RP). Sequence comparisons, Y2H and BiFC studies revealed that some are likely assembly chaperones, with several CP factors harboring the signature C-terminal HbYX motif that allows their association with the  $\alpha$ -subunit ring. Several of the RP-specific factors appear to be orthologs of the chaperones Nas2, Nas6, Hsm3 and Ecm29. Whereas yeast assembles only a single particle type, mammals can assemble alternate proteasomes by replacing individual subunits with distinct isoforms (e.g., immunoproteasomes). In plants, most 26S proteasome

subunits are encoded by paralogous genes with sufficient divergence to suggest that plants also accumulate a collection of particles. However, proteomic analysis of proteasomes selectively enriched using paralog-specific tags strongly imply that although plants possess this genetic diversity, the incorporation of these paralogs appears random, and is mainly influenced by the differential expression of the corresponding genes. Taken together, these proteomic studies provide the first insights into plant proteasome assembly and diversity, and identify factors that build the CP and RP subcomplexes and finally the 26S holo-particle.

## Chapter 1

### THE UBIQUITIN 26S PROTEASOME SYSTEM

#### 1.1. Introduction to the Ubiquitin 26S Proteasome System

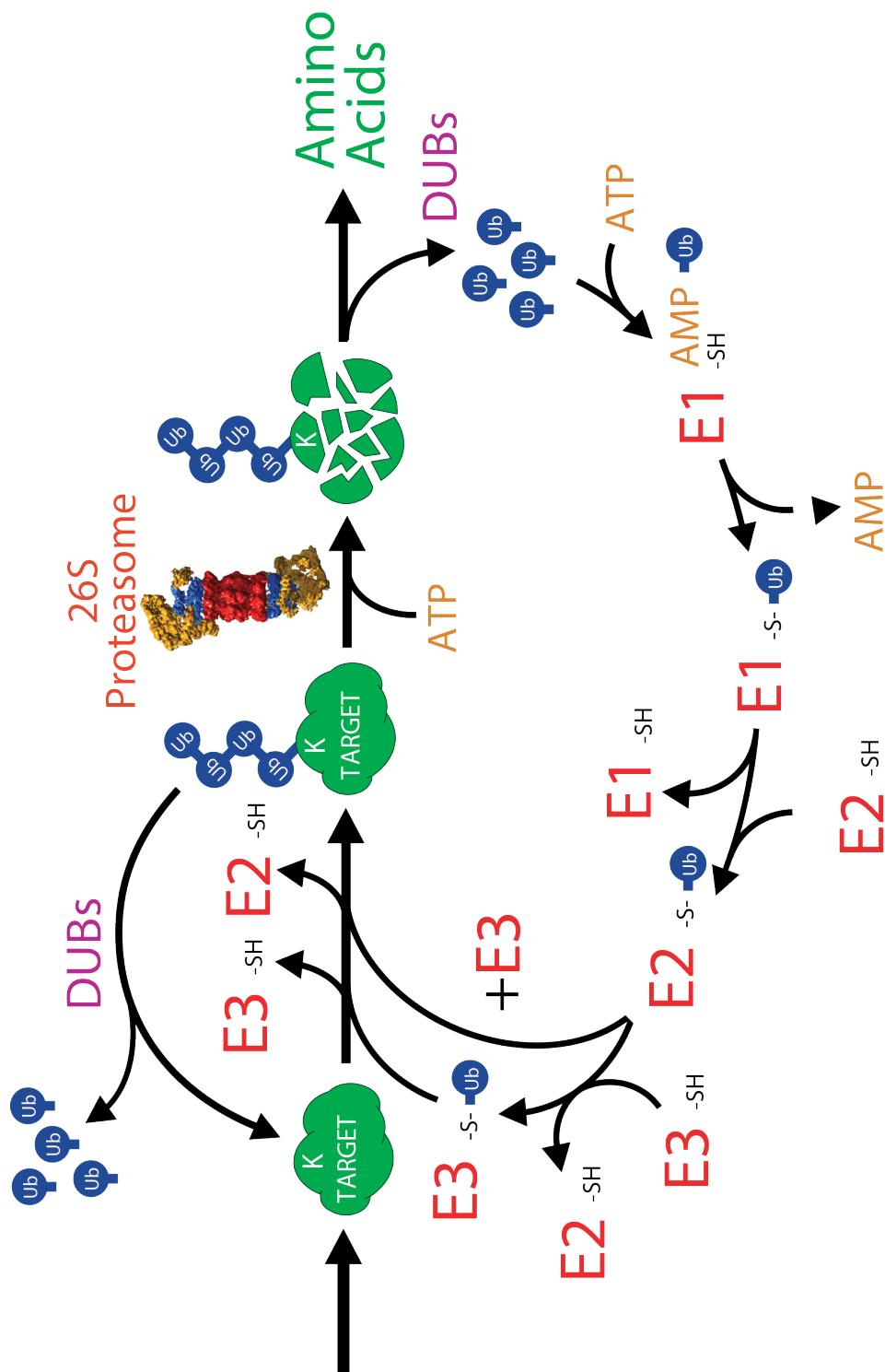
Selective proteolysis in plants plays a critical role in both regulating growth and development, and maintaining cellular homeostasis (Nelson et al., 2014; Smalle and Vierstra, 2004; Vierstra, 1993, 2009). One of the principle pathways for protein degradation in plants and other eukaryotes is the ubiquitin-26S proteasome system (UPS), which involves the attachment of polyubiquitin chains to target proteins followed by their recognition and degradation by the 26S proteasome, an exquisitely designed proteolytic machine (Bhattacharyya et al., 2014; Finley, 2009; Vierstra, 2009). The UPS is highly conserved across all eukaryotes; it was first elucidated by elegant work in rabbit reticulocyte lysates (Ciechanover et al., 1980a,b; Etlinger and Goldberg, 1977; Hershko et al., 1980; Wilkinson et al., 1980), and was subsequently identified in other animals, yeast and higher plants (Ciechanover et al., 1984; Finley et al., 1984, 1987; Glotzer et al., 1991; Hochstrasser et al., 1991; Shanklin et al., 1987). Ubiquitin conjugation to target proteins is accomplished through a highly polymorphic, ATP-dependent cascade involving the sequential action of three enzyme classes, termed the E1 ubiquitin-activating enzymes, E2 ubiquitin-conjugating en-

zymes, and E3 ubiquitin-protein ligases (Berndsen and Wolberger, 2014; Smalle and Vierstra, 2004; Vierstra, 2009). Selectivity in ubiquitylation is driven by the E3 family, which has dramatically expanded during plant evolution to include well over a thousand variants in *Arabidopsis thaliana* and other plant species (Hua et al., 2013; Hua and Vierstra, 2011). Through this myriad of E3s combined with the 26S proteasome, plants precisely control the levels of many key intracellular regulators that impact most, if not all, aspects of plant biology (Kim et al., 2013; Vierstra, 2009).

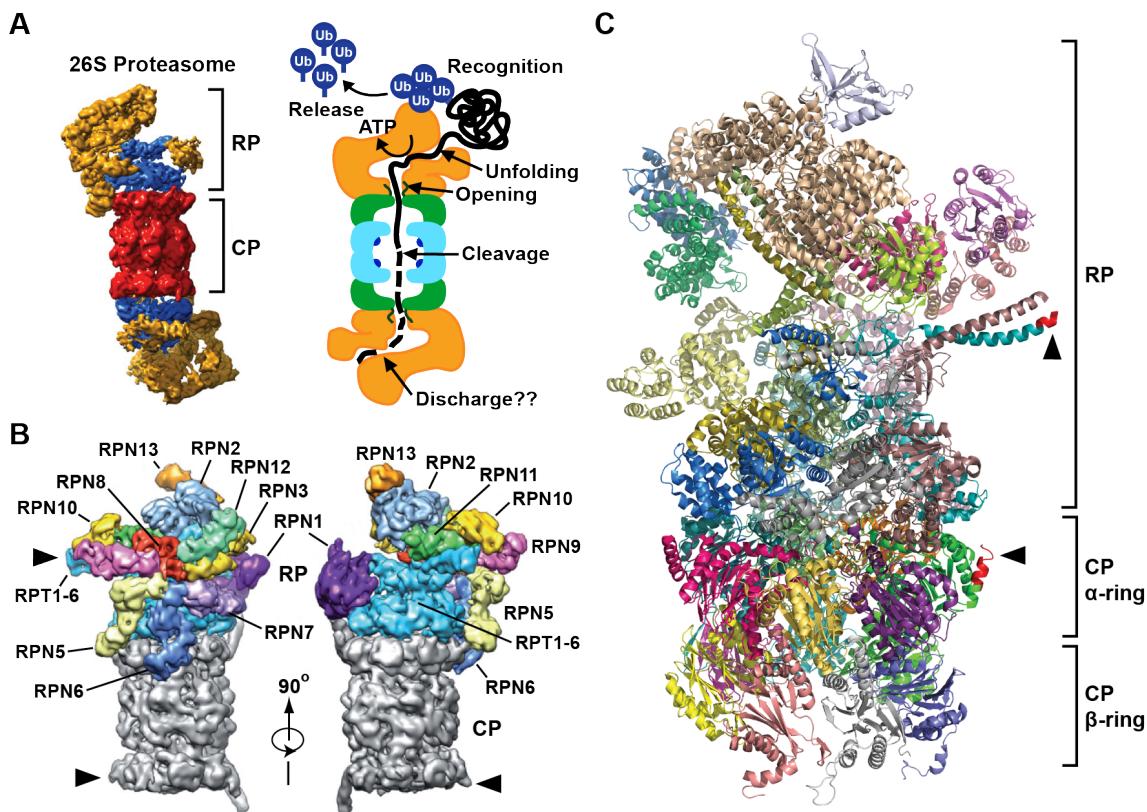
This process is illustrated in Figure 1.1.

## 1.2. The 26S Proteasome

The 26S proteasome is a 2.5 MDa particle located in the cytosol and nucleus of eukaryotic cells. It is composed of two functionally distinct sub-complexes; the 20S core protease (CP) that houses the proteolytic active sites, and the 19S regulatory particle (RP) that recognizes appropriate substrates (Figures 1.2 A and B; (Bhattacharyya et al., 2014; Finley, 2009; Lander et al., 2012; Lasker et al., 2012; Undvorben et al., 2014)). The CP has a barrel shape generated by four stacked hetero-heptameric rings, which contain seven  $\alpha$ -subunits or seven  $\beta$ -subunits (termed PAA-PAG and PBA-PBG, respectively, in *Arabidopsis*) in an  $\alpha$ 1-7/ $\beta$ 1-7/ $\beta$ 1-7/ $\alpha$ 1-7 configuration. Upon assembly, a central chamber is formed at the  $\beta$ -ring interface



**Figure 1.1: The Ubiquitin 26S Proteasome System.** Target proteins are covalently modified by ubiquitin an ATP-dependent E1 activating, E2 conjugating, E3 ligase enzymatic cascade. Once a target protein becomes polyubiquitylated, typically with a K48 linkage, it is efficiently recognized and subsequently degraded by the 26S proteasome. Ubiquitin is released from the target protein by various de-ubiquitylating enzymes and can then enter the cycle again to covalently modify other substrates.



**Figure 1.2: Structure of the 26S proteasome.** (A) Schematic representation of the 26S proteasome, with a 3-D structure as determined by electron microscopy (EM) shown on the left, and a cartoon representation of the holoprotease shown on the right. Within the EM structure, the CP is shown in red, the RP base is shown in blue, and the RP lid is shown in yellow. Specific functions within the CP and RP are shown on the right. The EM structure is modified from reference (Lasker et al., 2012) (B) A detailed view of the subunit architecture of the 26S proteasome RP. The CP is shown in grey, the RPT ring is shown in blue, and all additional RPN subunits are shown in different colors, with their identity labelled. The positions of the FLAG tags on PAG1 and RPT4 are indicated by red arrowheads. These structures are modified from reference (Lander et al., 2012). (C) A structural model of the 26S proteasome from yeast at sub-atomic resolution modified from PDB ID 4CR2 (Beck et al., 2012). The RP subunits, as well as the CP  $\alpha$  and  $\beta$  rings are shown. Highlighted in red, and indicated by black arrowheads, are the positions where FLAG affinity tags have been successfully used to enrich for *Arabidopsis* 26S proteasomes. The affinity purification of the RP in this thesis exploits the tag shown in the regulatory particle.

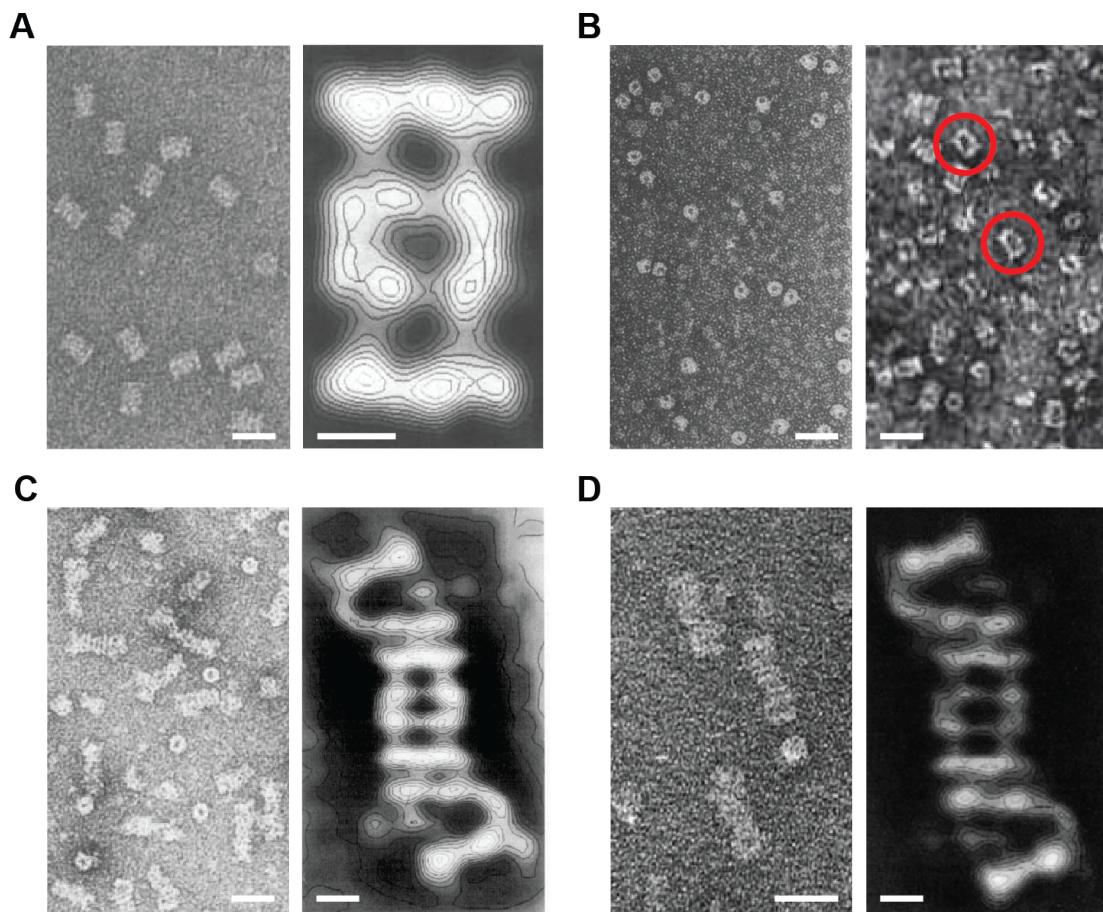
that houses six peptidase catalytic sites provided by the  $\beta$ 1 (PBA),  $\beta$ 2 (PBB), and  $\beta$ 5 (PBE) subunits (Arendt and Hochstrasser, 1997; Heinemeyer et al., 1997). The active sites involve a catalytic triad, one residue of which is an N-terminal threonine that becomes exposed during CP assembly. Collectively these peptidases can cleave a broad range of protein sequences (Arendt and Hochstrasser, 1997; Groll et al., 1999). The  $\alpha$ -rings create two antechambers with narrow opposing axial pores that are gated by extensions at the N-terminus of several subunits (Groll et al., 2000; Ruschak et al., 2010). Through this distinctive architecture, the CP acts as a self-compartmentalized protease that will only degrade polypeptides that are deliberately recognized, unfolded, and imported into the  $\beta$ -ring chamber. The CP is capped at one or both ends by the RP, which sits on top of the axial pores. The RP provides activities for recognition of ubiquitylated proteins, substrate unfolding and import, and release of the ubiquitin moieties before substrate degradation. Its binding to the CP is stabilized by ATP, which is thus a necessary ingredient for purifying intact 26S proteasomes. The RP itself consists of two sub-complexes; the base, which contains a hexameric ring of AAA-ATPases (RPT1-6) plus two non-ATPase subunits, RPN1 and RPN2; and the lid, which is composed of an additional 11 non-ATPase subunits, RPN3, RPN5-13 and DSS1/SEM1 (Figures 1.2 B and C; (Bhattacharyya et al., 2014; Book et al., 2010; Finley, 2009; Glickman et al., 1998b;

Russell et al., 2013). This lid/base demarcation was first revealed by the absence of lid subunits in proteasomes isolated from a  $\Delta rpn10$  yeast deletion strain, and it was hence thought that RPN10 helps enforce binding of the lid to the base (Glickman et al., 1998a). However, more recent structural studies have demonstrated that RPN10 has a more indirect stabilizing role via its interaction with RPN9 (Lander et al., 2012). The ring of RPT subunits in the base promotes substrate unfolding through ATP hydrolysis, and gates the  $\alpha$ -ring axial pores through repositioning of the CP  $\alpha$ -subunit extensions (Köhler et al., 2001; Rabl et al., 2008; Smith et al., 2005). The N-terminal regions from proximal RPT pairs intertwine to create three spokes onto which most RPN subunits are scaffolded (Figure 1.2 C; (Beck et al., 2012)). The RPN6 subunit acts as a molecular clamp to tether the RP onto the CP (Pathare et al., 2012), while RPN11 is a metalloprotease that uses a zinc-coordinated active site to release the ubiquitin moieties isopeptide-linked to substrates (Verma et al., 2002; Worden et al., 2014). Through RPN11 and other loosely associated deubiquitylating enzymes such as UBP6/USP14 (Hanna et al., 2006; Sakata et al., 2011), bound ubiquitins are actively recycled. Substrate selection by the 26S proteasome is dictated by several ubiquitin receptors intrinsic to the RP lid, including RPN10, RPN13, and DSS1/SEM1 (Fatimababy et al., 2010; Finley, 2009; Lin et al., 2011; Paraskevopoulos et al., 2014; Sakata et al., 2012; Van Nocker et al., 1996), and

possibly RPN1 in the base (Elsasser et al., 2002). RPN10 binds ubiquitin via defined ubiquitin-interacting motifs (UIMs), of which yeast, human and *Arabidopsis* RPN10 contain 1, 2 and 3 in tandem, respectively (Fatimababy et al., 2010; Finley, 2009; Fu et al., 1998; Lin et al., 2011; Van Nocker et al., 1996). By contrast, RPN13 binds ubiquitin via a pleckstrin-like receptor for ubiquitin (PRU) domain, which is structurally distinct from UIMs but binds to the same hydrophobic patch on ubiquitin (Husnjak et al., 2008; Schreiner et al., 2008). More recently, DSS1/SEM1 was also found to be a proteasomal ubiquitin receptor (Paraskevopoulos et al., 2014). It had previously resisted identification due to both its small size, which prevented visualization by standard protein stains following sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and its paucity of lysine and arginine residues, which complicated detection by conventional mass spectrometric methods. Only with the use of top-down mass spectrometry of 26S proteasome complexes was DSS1/SEM1 first detected in intact 26S proteasomes from *Arabidopsis* (Russell et al., 2013). In addition to these core ubiquitin receptors, there are several extra-proteasomal ubiquitin-binding proteins that shuttle ubiquitylated cargo to the RP. They work by virtue of ubiquitin-associated (UBA) domains that bind ubiquitin, combined with a ubiquitin-like (UBL) domain that interacts with the intrinsic ubiquitin receptors such as RPN10. Important shuttle factors in plants include RAD23, DSK2, and DDI1

(Farmer et al., 2010; Fatimababy et al., 2010; Finley, 2009; Lin et al., 2011), though many other ubiquitin-binding proteins are known in other species (Husnjak and Dikic, 2012). Numerous other factors also associate sub-stoichiometrically with the mature CP and RP sub-complexes, including deubiquitylating enzymes, several E3 ligases and protein kinases, and a collection of protein folding chaperones (Besche et al., 2014; Book et al., 2010; Leggett et al., 2002; Xie and Varshavsky, 2000). Not surprisingly given its intricate architecture, construction of the 26S proteasome requires a large collection of assembly factors that work in synchrony. Included are chaperones required for the correctly ordered assembly of the  $\alpha$ - and  $\beta$ -rings of the CP and the RPT ring of the RP, which in yeast involve the Pba1/2 and Pba3/4 heterodimers for the CP (Kusmierczyk et al., 2008; Le Tallec et al., 2007; Tomko and Hochstrasser, 2013), and Nas2, Nas6, Hsm3 and Rpn14 for the RP (Funakoshi et al., 2009; Roelofs et al., 2009; Saeki et al., 2009; Tomko and Hochstrasser, 2013). Additional chaperones then mediate assembly of the final particle. UMP1 is required to connect the two  $\alpha/\beta$  half-barrels to generate the complete CP. Once its job is finished UMP1 is degraded, thus becoming the first proteolytic substrate of the fully assembled CP (Ramos et al., 1998). ECM29 stabilizes the association of assembled CP and RP and provides a final quality control checkpoint for mature 26S proteasomes (Besche et al., 2014; Lehmann et al., 2010). Finally, in some situa-

tions, the RP is replaced entirely by alternate capping particles such as PA200 (also known as Blm10) or CDC48 (Barthelme and Sauer, 2012; Book et al., 2010; Schmidt et al., 2005). The functions of these caps are not yet clear, but recent proposals for PA200 have it participating in 26S proteasome assembly, helping shuttle proteasomes into the nucleus, and/or generating a ubiquitin-independent proteasome containing CP and PA200 only (Dange et al., 2011; Sadre-Bazzaz et al., 2010; Weber-russ et al., 2013). Even before the realization that the 26S proteasome is a protease, sub-particles of the complexes were described. The first reports of proteasomes used avian erythroblast preparations enriched by differential ultracentrifugation followed by fractionation through a sucrose gradient (Schmid et al., 1984). These 20S fractions isolated in the absence of added ATP were found to inhibit mRNA translation in a cell-free system, leading to early proposals that the identified complex repressed gene expression through a cryptic ribonuclease activity. This lead to the particle initially being named the “prosome” (Kremp et al., 1986; Schmid et al., 1984). Subsequent analyses of these preparations by SDS-PAGE and electron microscopy revealed the signature ladder of  $\alpha$ - and  $\beta$ -subunits at 20-35 kDa, as well as their barrel-like architecture (Figures 1.3 A and B; (Baumeister et al., 1988; Kremp et al., 1986; Schmid et al., 1984)). Purification of the 20S fraction from HeLa cells followed by SDS-PAGE also gave rise to this stereotypical protein banding



**Figure 1.3: Electron microscopy images of the 20S and 26S proteasomes from mammals and plants.** (A) Images of 20S proteasomes purified from rat skeletal muscle. On the left is an electron micrograph of the 20S particles negatively stained with sodium phosphotungstate, while on the right is a close-up image with overlaid contour plots generated by correlation averaging of approximately 300 individual images negatively stained with ammonium molybdate. (B) Images of the first 20S proteasomes purified from different plant species. On the left are proteasomes isolated from tobacco leaves, while on the right are proteasomes from potato tubers, both negatively stained with uranyl acetate. The typical barrel-shaped structures are indicated with red circles. (C) Images of 26S proteasomes purified from rat liver. On the left is an electron micrograph of the 26S particles negatively stained with uranyl acetate, while on the right is a close-up image with overlaid contour plots generated by correlation averaging of 215 individual images. (D) Images of 26S proteasomes purified from spinach leaves. On the left is an electron micrograph of the 26S particles negatively stained with uranyl acetate, while on the right is a close-up image with overlaid contour plots generated by correlation averaging of 450 individual images. In all cases, scale bars represent 25 nm for the electron micrograph images and 5 nm for close-up images generated by averaging. The images were modified from references (Baumeister et al., 1988; Fujinami et al., 1994; Kremp et al., 1986; Schliephacke et al., 1991; Yoshimura et al., 1993).

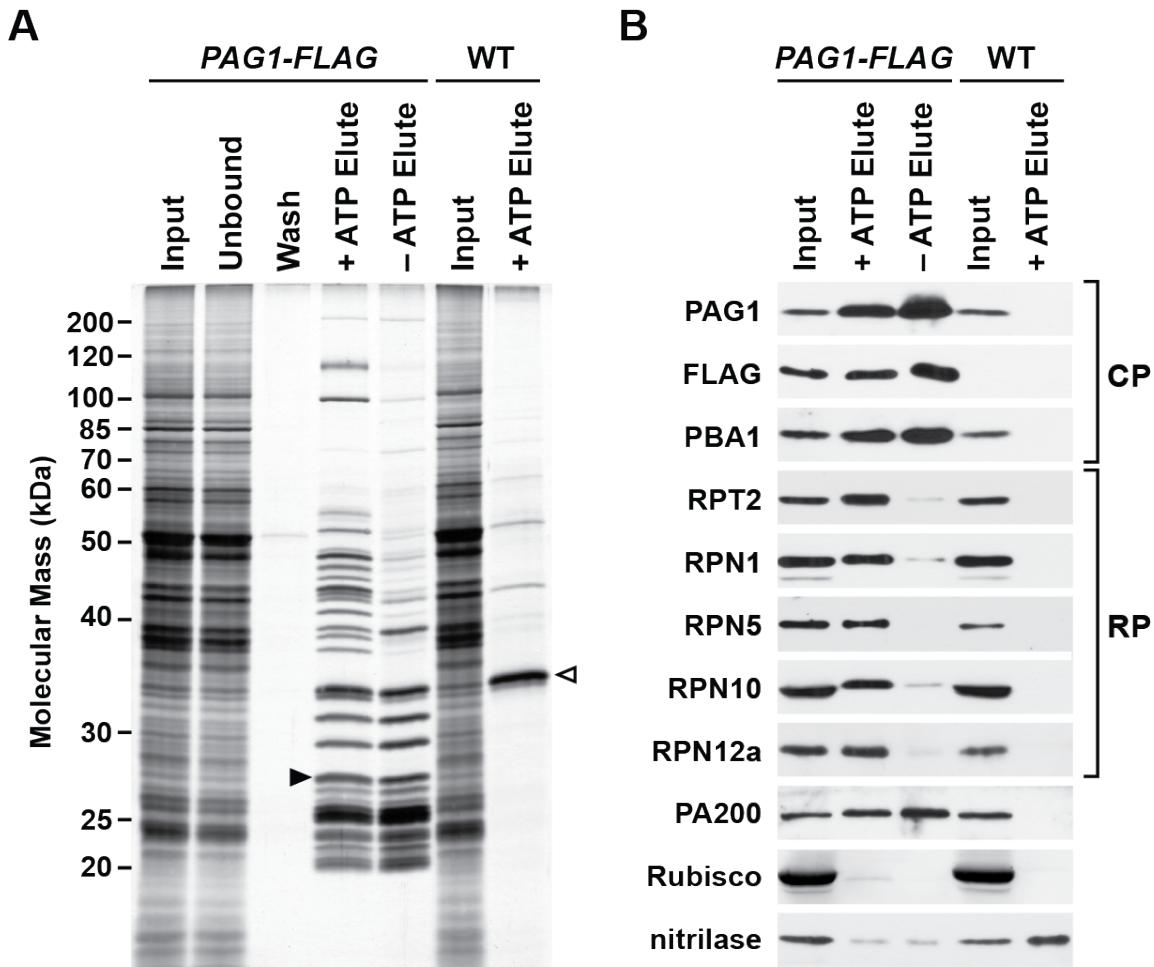
pattern and shape (Schmid et al., 1984), and this was followed shortly thereafter by the first description of plant prosomes, purified from tobacco leaf extracts using similar sedimentation protocols in ATP-free buffers (Kremp et al., 1986). In these later cases, the purified preparations had strong peptidase activity but little to no RNase activity, thus leading to the conclusion that the CP is actually a protease. Once its true function in protein turnover was confirmed, the moniker for the particle was changed to 'proteasome' (Arrigo et al., 1988). Subsequently, the 20S particle was purified from other plant tissues, including dry pea seeds, potato tubers, mung bean seedlings, and leaves from both spinach and wheat (Murray et al., 1997; Ozaki et al., 1992; Schliephacke et al., 1991; Skoda and Malek, 1992). These purifications were typically performed using sequential anion exchange and size-exclusion chromatography steps in the absence of ATP, hence only the CP was isolated. Their remarkable similarity in protein composition and structure, as observed by SDS-PAGE and electron microscopy, respectively, coupled with the fact that several of the plant subunits cross-reacted with antibodies against their yeast, human, rat and Xenopus counterparts, strongly implied that the CP was conserved and widely distributed among eukaryotes (Schliephacke et al., 1991). The complete 26S proteasome (i.e. the CP capped at one or both ends by the RP) was subsequently discovered by the purification of ubiquitin conjugate-degrading activity

from rabbit reticulocytes (Hough et al., 1986). While it had been well established that major catabolic processes in animal cells involved the ATP-dependent proteolysis of selective substrates (Etlinger and Goldberg, 1977), the enzyme(s) responsible for this activity had yet to be identified. Taking advantage of the new ability to synthesize ubiquitylated substrates such as  $^{125}\text{I}$ -labelled ubiquitin-lysozyme conjugates (Hough and Rechsteiner, 1986), a protocol was developed to purify the responsible ATP-dependent protease. Through a series of anion exchange and size exclusion chromatography steps followed by glycerol gradient sedimentation, all of which were performed in ATP-containing buffers, the responsible activity was isolated (Hough et al., 1986, 1987). The active enzyme turned out to be the 20S proteasome (i.e. the CP) along with a number of additional polypeptides which together formed a 26S particle, thus providing the first direct link between ubiquitylation and a protease (Ganoth et al., 1988; Hough et al., 1987; Waxman et al., 1987). SDS-PAGE analysis of these preparations identified a host of new polypeptides in the 35-100 kDa range in addition to the known CP subunits, which were later shown to comprise a second stable complex, the RP. Shortly thereafter, the RP was demonstrated to have ATPase activities attributable to the RPT subunits, which help in substrate unfolding and maintaining CP-RP association (Armon et al., 1990). Electron microscopic images of the full 26S particle then revealed its diagnostic

quaternary structure in which the CP is capped by one or two RPs which sit over the axial pores for substrate entry (Figure 1.3 C; (Peters et al., 1991; Yoshimura et al., 1993)). The existence of a similar 26S proteasome in plants was initially implied by the detection of an ATP-dependent activity in oat and wheat germ extracts capable of degrading ubiquitylated proteins (Hatfield and Vierstra, 1989; Vierstra and Sullivan, 1988). This was followed some years later by the first isolation of a complete plant 26S proteasome holocomplex from spinach leaves (Fujinami et al., 1994). As with the mammalian forms, purification was achieved by anion exchange and size exclusion chromatography, followed by glycerol gradient centrifugation, all in the presence of ATP to stabilize the CP-RP association. These spinach preparations were, like their rabbit reticulocyte counterparts, able to rapidly degrade ubiquitylated substrates in an ATP-dependent manner, and further analysis by native-PAGE, SDS-PAGE and electron microscopy revealed the complete subunit composition and “caterpillar-like” structure of the plant particle (Figure 1.3 D; (Fujinami et al., 1994)). Similar purifications were successful using rice suspension culture cells and garlic cloves (Malik et al., 2004; Yanagawa et al., 1999), which were accompanied by the first demonstrations that proteasome inhibitors designed for their mammalian counterparts were effective with the plant particles, suggesting very similar enzymatic mechanisms (Ozaki et al., 1992; Woffenden et al., 1998).

Despite its prevalence as a genetic model, purification of the 26S proteasome from the flowering plant *Arabidopsis thaliana* was not reported until several years after other plant species (Yang et al., 2004). First protocols involved differential PEG precipitation followed by anion exchange and size exclusion chromatography, with the latter exploiting the large size of the holoprotease. More recently, an improved one-step affinity method was developed (Book et al., 2010), based on the strategies that had been successfully employed in yeast (Leggett et al., 2005). Here, epitope-tagged proteasomes were generated by genetically replacing the subunit PAG1 with a variant bearing a C-terminal tag; this tagged particle could then be purified with appropriate affinity matrices. This approach enables rapid and robust purification of the whole 26S proteasome complex when performed in the presence of ATP, or enables purification of the CP sub-particle when performed in the absence of ATP (Book et al., 2010). The *Arabidopsis* 26S proteasome exists in planta as a diverse array of complexes containing multiple subunit isoforms and interacting proteins (Book et al., 2010; Fu et al., 1999; Yang et al., 2004). To facilitate biochemical analysis of the plant particle, we developed a rapid and robust affinity purification protocol that enables isolation of intact 26S proteasomes, and the individual CP and RP sub-complexes, by genetically replacing individual subunits with FLAG-tagged versions (Book et al., 2010). Such a strategy was based on a similar approach used

successfully with yeast, where the proteasome subunits Pre1, Rpt1 and Rpn11 were appended with either FLAG or Protein A tags to permit effective affinity enrichment (Leggett et al., 2005). Using the recently described structures of the 26S proteasome (Bhattacharyya et al., 2014; Lander et al., 2012; Lasker et al., 2012), we identified subunits in the CP (PAG1) and as demonstrated in this thesis RP (RPT4) which had solvent exposed N- or C-termini that were potentially appropriate for appending the epitope tag (see Figures 1.2 B and C). As mentioned above, the affinity method has considerable advantages compared to previous conventional chromatographic approaches (Yang et al., 2004) as it is both faster and more reliable, produces higher yields per gram of tissue ( 6  $\mu$ g/g), and allows purification of the CP and RP separately by omitting ATP from the buffers and/or performing a high salt wash step prior to elution. Additionally, it also avoids the harsh buffer conditions necessary for conventional purification, which has allowed the identification of less tightly bound core and accessory components, such as various CP and RP assembly chaperones, the ubiquitin receptors RPN13 and DSS1/SEM1, and the alternate capping particle PA200 (Book et al., 2010; Russell et al., 2013). This milder more rapid technique also prevents breakdown of some subunits, in particular RPN10, which is sensitive to post-homogenization proteolysis (Yang et al., 2004). One caveat is that the epitope tag, given its exposed position and flexible structure, might be sen-



**Figure 1.4: Affinity purification of 26S proteasomes from *pag1-1 PAG1-FLAG* plants.** (A) SDS-PAGE analysis of the affinity purification steps. Total protein extracts from 10 day old wild-type (WT) and *pag1-1* *PAG1-FLAG* plants were incubated with anti-FLAG affinity resin, washed, and competitively eluted with the FLAG peptide. The procedure was performed in the presence or absence of ATP, and the input, unbound, washed and eluted fractions were subjected to SDS-PAGE and the gel stained for protein with silver. The black arrowhead indicates the *PAG1-FLAG* protein, while the open arrowhead identifies nitrilase, which is non-specifically enriched during the purification. (B) Immunoblot detection of various 26S proteasome subunits in the affinity-purified preparations shown in A. Subunits tested include the CP subunits PAG1 and PBA1, the RP subunits RPT2, RPN1, RPN5, RPN10 and RPN12a, and the alternate capping particle PA200. Other proteins tested include the Rubisco small subunit and nitrilase. This figure was modified from reference (Book et al., 2010).

sitive to proteolytic cleavage following tissue homogenization. For the PAG1-FLAG protocol, chymostatin was found to effectively block the interfering protease (Book et al., 2010). An example of such preparations analyzed by SDS-PAGE followed by immunoblotting with antibodies against several proteasome subunits, are shown in Figures 1.4 A and B, respectively. Keep in mind that ATP should be included in all purification steps if the final desired outcome is the fully assembled 26S particle.

### **1.3. 26S Proteasome Assembly**

### **1.4. Proteasome Post-Translational Modification**

### **1.5. Proteasome Degredation**

### **1.6. Proteasome Interacting Proteins**

### **1.7. References**

**Arendt, C.S. and Hochstrasser, M.** (1997). Identification of the yeast 20s proteasome catalytic centers and subunit interactions required for active site formation. Proceedings of the National Academy of Sciences of the United States of America **94**(14):7156–7161.

**Armon, T., Ganot, D., and Hershko, A.** (1990). Assembly of the 26s complex that

degrades proteins ligated to ubiquitin is accompanied by the formation of atpase activity. *Journal of Biological Chemistry* **265**(34):20723–20726.

**Arrigo, A.P., Tanaka, K., Goldberg, A.L., and Welch, W.J.** (1988). Identity of the 19s prosome particle with the large multi-functional protease complex of mammalian cells (the proteasome). *Nature* **331**(6152):192–194.

**Barthelme, D. and Sauer, R.T.** (2012). Identification of the cdc48-20s proteasome as an ancient aaa proteolytic machine. *Science* **337**(6096):843–846.

**Baumeister, W., Dahlmann, B., Hegerl, R., Kopp, F., KÄ¼hn, L., and Pfeifer, G.** (1988). Electron microscopy and image analysis of the multi-catalytic proteinase. *FEBS Letters* **241**(1-2):239–245.

**Beck, F., Unverdorben, P., Bohn, S., Schweitzer, A., Pfeifer, G., Sakata, E., Nickell, S., Plitzko, J.M., Villa, E., Baumeister, W., and FÄ¶rster, F.** (2012). Near-atomic resolution structural model of the yeast 26s proteasome. *Proceedings of the National Academy of Sciences of the United States of America* **109**(37):14870–14875.

**Berndsen, C.E. and Wolberger, C.** (2014). New insights into ubiquitin e3 ligase mechanisms. *Nature Structural & Molecular Biology* **21**(4):301–307.

- Besche, H.C., Sha, Z., Kukushkin, N.V., Peth, A., Hock, E.M., Kim, W., Gygi, S.P., Gutierrez, J.A., Liao, H., Dick, L.R., and Goldberg, A.L.** (2014). Auto-ubiquitination of the 26s proteasome on rpn13 regulates breakdown of ubiquitin conjugates. *EMBO Journal* **33**(10):1159–1176.
- Bhattacharyya, S., Yu, H., Mim, C., and Matouschek, A.** (2014). Regulated protein turnover: snapshots of the proteasome in action. *Nature Reviews Molecular Cell Biology* **15**(2):122–133.
- Book, A.J., Gladman, N.P., Lee, S.S., Scalf, M., Smith, L.M., and Vierstra, R.D.** (2010). Affinity purification of the arabidopsis 26s proteasome reveals a diverse array of plant proteolytic complexes. *Journal of Biological Chemistry* **285**(33):25554–25569.
- Ciechanover, A., Elias, S., Heller, H., Ferber, S., and Hershko, A.** (1980a). Characterization of the heat-stable polypeptide of the atp-dependent proteolytic system from reticulocytes. *Journal of Biological Chemistry* **255**(16):7525–8.
- Ciechanover, A., Finley, D., and Varshavsky, A.** (1984). Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. *Cell* **37**(1):57–66.

- Ciechanover, A., Heller, H., Elias, S., Haas, A.L., and Hershko, A.** (1980b). Atp-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proceedings of the National Academy of Sciences of the United States of America* **77**(3):1365–1368.
- Dange, T., Smith, D.M., Noy, T., Rommel, P.C., Jurzitzza, L., Cordero, R.J., Legendre, A., Finley, D., Goldberg, A.L., and Schmidt, M.** (2011). The blm10 protein promotes proteasomal substrate turnover by an active gating mechanism. *Journal of Biological Chemistry* **286**(50):42830–42839.
- Elsasser, S., Gali, R.R., Schwickart, M., Larsen, C.N., Leggett, D.S., MÃ¼ller, B., Feng, M.T., TÃ¼bing, F., Dittmar, G.A., and Finley, D.** (2002). Proteasome subunit rpn1 binds ubiquitin-like protein domains. *Nature Cell Biology* **4**(9):725–730.
- Etlinger, J.D. and Goldberg, A.L.** (1977). A soluble, atp-dependent proteolytic system responsible for degradation of abnormal proteins in reticulocytes. *Proceedings of the National Academy of Sciences of the United States of America* **74**(1):54–58.
- Farmer, L.M., Book, A.J., Lee, K.H., Lin, Y.L., Fu, H., and Vierstra, R.D.** (2010).

The rad23 family provides an essential connection between the 26s proteasome and ubiquitylated proteins in arabidopsis. *Plant Cell* **22**(1):124–142.

**Fatimababy, A.S., Lin, Y.L., Usharani, R., Radjacompare, R., Wang, H.T., Tsai, H.L., Lee, Y., and Fu, H.** (2010). Cross-species divergence of the major recognition pathways of ubiquitylated substrates for ubiquitin/26s proteasome-mediated proteolysis. *FEBS Journal* **277**(3):796–816.

**Finley, D.** (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual Review of Biochemistry* **78**:477–513.

**Finley, D., Ciechanover, A., and Varshavsky, A.** (1984). Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. *Cell* **37**(1):43–55.

**Finley, D., Ozkaynak, E., and Varshavsky, A.** (1987). The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses. *Cell* **48**(6):1035–1046.

**Fu, H., Girod, P.A., Doelling, J.H., Van Nocker, S., Hochstrasser, M., Finley, D., and Vierstra, R.D.** (1999). Structural and functional analyses of the 26s proteasome subunits from plants. *Molecular Biology Reports* **26**(1-2):137–146.

- Fu, H., Sadis, S., Rubin, D.M., Glickman, M.H., Van Nocker, S., Finley, D., and Vierstra, R.D.** (1998). Multi-ubiquitin chain binding and protein degradation are mediated by distinct domains within the 26s proteasome subunit mcb1. *Journal of Biological Chemistry* **273**(4):1970–1981.
- Fujinami, K., Tanahashi, N., Tanaka, K., Ichihara, A., Cejka, Z., Baumeister, W., Miyawaki, M., Sato, T., and Nakagawa, H.** (1994). Purification and characterization of the 26s proteasome from spinach leaves. *Journal of Biological Chemistry* **269**(41):25905–25910.
- Funakoshi, M., Tomko, R.J., Kobayashi, H., and Hochstrasser, M.** (2009). Multiple assembly chaperones govern biogenesis of the proteasome regulatory particle base. *Cell* **137**(5):887–899.
- Ganoth, D., Leshinsky, E., Eytan, E., and Hershko, A.** (1988). A multi-component system that degrades proteins conjugated to ubiquitin: resolution of factors and evidence for atp-dependent complex formation. *Journal of Biological Chemistry* **263**(25):12412–12419.
- Glickman, M.H., Rubin, D.M., Coux, O., Wefes, I., Pfeifer, G., Cjeka, Z., Baumeister, W., Fried, V.A., and Finley, D.** (1998a). A subcomplex of the pro-

teasome regulatory particle required for ubiquitin-conjugate degradation and related to the cop9-signalosome and eif3. *Cell* **94**(5):615–623.

**Glickman, M.H., Rubin, D.M., Fried, V.A., and Finley, D.** (1998b). The regulatory particle of the *saccharomyces cerevisiae* proteasome. *Molecular and Cellular Biology* **18**(6):3149–3162.

**Glotzer, M., Murray, A.W., and Kirschner, M.W.** (1991). Cyclin is degraded by the ubiquitin pathway. *Nature* **349**(6305):132–138.

**Groll, M., Bajorek, M., Kähler, A., Moroder, L., Rubin, D.M., Huber, R., Glickman, M.H., and Finley, D.** (2000). A gated channel into the proteasome core particle. *Nature Structural Biology* **7**(11):1062–1067.

**Groll, M., Heinemeyer, W., Jäger, S., Ullrich, T., Bochtler, M., Wolf, D.H., and Huber, R.** (1999). The catalytic sites of 20s proteasomes and their role in subunit maturation: a mutational and crystallographic study. *Proceedings of the National Academy of Sciences of the United States of America* **96**(20):10976–10983.

**Hanna, J., Hathaway, N.A., Tone, Y., Crosas, B., Elsasser, S., Kirkpatrick, D.S., Leggett, D.S., Gygi, S.P., King, R.W., and Finley, D.** (2006). The deubiquitinating enzyme ubp6 functions non-catalytically to delay proteasomal degradation. *Cell* **127**(1):99–111.

**Hatfield, P.M. and Vierstra, R.D.** (1989). Ubiquitin-dependent proteolytic pathway in wheat germ: isolation of multiple forms of the e1 ubiquitin-activating enzyme. *Biochemistry* **28**(2):735–742.

**Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D.H.** (1997). The active sites of the eukaryotic 20s proteasome and their involvement in subunit precursor processing. *Journal of Biological Chemistry* **272**(40):25200–25209.

**Hershko, A., Ciechanover, A., Heller, H., Haas, A.L., and Rose, I.A.** (1980). Proposed role of atp in protein breakdown: conjugation of proteins with multiple chains of the polypeptide of atp-dependent proteolysis. *Proceedings of the National Academy of Sciences of the United States of America* **77**(4):1783–1786.

**Hochstrasser, M., Ellison, M.J., Chau, V., and Varshavsky, A.** (1991). The short-lived mat $\hat{I}\pm 2$  transcriptional regulator is ubiquitinated in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **88**(11):4606–4610.

**Hough, R., Pratt, G., and Rechsteiner, M.** (1986). Ubiquitin-lysozyme conjugates: identification and characterization of an atp-dependent protease from rabbit reticulocyte lysates. *Journal of Biological Chemistry* **261**(5):2400–2408.

**Hough, R., Pratt, G., and Rechsteiner, M.** (1987). Purification of two high molecular

weight proteases from rabbit reticulocyte lysate. *Journal of Biological Chemistry* **262**(17):8303–8313.

**Hough, R. and Rechsteiner, M.** (1986). Ubiquitin-lysozyme conjugates: purification and susceptibility to proteolysis. *Journal of Biological Chemistry* **261**(5):2391–2399.

**Hua, Z., Pool, J.E., Schmitz, R.J., Schultz, M.D., Shiu, S.H., Ecker, J.R., and Vierstra, R.D.** (2013). Epigenomic programming contributes to the genomic drift evolution of the f-box protein superfamily in arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **110**(42):16927–16932.

**Hua, Z. and Vierstra, R.D.** (2011). The cullin-ring ubiquitin-protein ligases. *Annual Review of Plant Biology* **62**:299–334.

**Husnjak, K. and Dikic, I.** (2012). Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annual Review of Biochemistry* **81**:291–322.

**Husnjak, K., Elsasser, S., Zhang, N., Chen, X., Randles, L., Shi, Y., Hofmann, K., Walters, K.J., Finley, D., and Dikic, I.** (2008). Proteasome subunit rpn13 is a novel ubiquitin receptor. *Nature* **453**(7194):481–488.

**KÃ¶hler, A., Cascio, P., Leggett, D.S., Woo, K.M., Goldberg, A.L., and Finley, D.** (2001). The axial channel of the proteasome core particle is gated by the rpt2

atpase and controls both substrate entry and product release. *Molecular Cell* **7**(6):1143–1152.

**Kim, D.Y., Scalf, M., Smith, L.M., and Vierstra, R.D.** (2013). Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in arabidopsis. *Plant Cell* **25**(5):1523–1540.

**Kremp, A., Schliephacke, M., Kull, U., and Schmid, H.P.** (1986). Prosomes exist in plant cells too. *Experimental Cell Research* **166**(2):553–557.

**Kusmierczyk, A.R., Kunjappu, M.J., Funakoshi, M., and Hochstrasser, M.** (2008). A multimeric assembly factor controls the formation of alternative 20s proteasomes. *Nature Structural & Molecular Biology* **15**(3):237–244.

**Lander, G.C., Estrin, E., Matyskiela, M.E., Bashore, C., Nogales, E., and Martin, A.** (2012). Complete subunit architecture of the proteasome regulatory particle. *Nature* **482**(7384):186–191.

**Lasker, K., FÄ¶rster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W.** (2012). Molecular architecture of the 26s proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences of the United States of America* **109**(5):1380–1387.

- Le Tallec, B., Barrault, M.B., Courbeyrette, R., GuÃ©rois, R., Marsolier-Kergoat, M.C., and Peyroche, A.** (2007). 20s proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and mammals. *Molecular Cell* **27**(4):660–674.
- Leggett, D.S., Glickman, M.H., and Finley, D.** (2005). Purification of proteasomes, proteasome sub-complexes, and proteasome-associated proteins from budding yeast. *Methods in Molecular Biology* **301**:57–70.
- Leggett, D.S., Hanna, J., Borodovsky, A., Crosas, B., Schmidt, M., Baker, R.T., Walz, T., Ploegh, H.L., and Finley, D.** (2002). Multiple associated proteins regulate proteasome structure and function. *Molecular Cell* **10**(3):495–507.
- Lehmann, A., Niewienda, A., Jechow, K., Janek, K., and Enenkel, C.** (2010). Ecm29 fulfills quality control functions in proteasome assembly. *Molecular Cell* **38**(6):879–888.
- Lin, Y.L., Sung, S.C., Tsai, H.L., Yu, T.T., Radjacommare, R., Usharani, R., Fatimababy, A.S., Lin, H.Y., Wang, Y.Y., and Fu, H.** (2011). The defective proteasome, not substrate recognition function, is responsible for the null phenotypes of the arabidopsis proteasome subunit rpn10. *Plant Cell* **23**(7):2754–2773.
- Malik, M.N., Spivack, W.D., Sheikh, A.M., and Fenko, M.D.** (2004). The 26s

proteasome in garlic (*Allium sativum*): purification and partial characterization. *Journal of Agricultural and Food Chemistry* **52**(11):3350–3355.

**Murray, P.F., Giordano, C.V., Passeron, S., and Barneix, A.J.** (1997). Purification and characterization of the 20s proteasome from wheat leaves. *Plant Science* **125**(2):127–136.

**Nelson, C.J., Li, L., and Millar, A.H.** (2014). Quantitative analysis of protein turnover in plants. *Proteomics* **14**(4-5):579–592.

**Ozaki, M., Fujinami, K., Tanaka, K., Amemiya, Y., Sato, T., Ogura, N., and Nakagawa, H.** (1992). Purification and initial characterization of the proteasome from the higher plant *Spinacia oleracea*. *Journal of Biological Chemistry* **267**(30):21678–21684.

**Paraskevopoulos, K., Kriegenburg, F., Tatham, M.H., RÃ¶sner, H.I., Medina, B., Larsen, I.B., Brandstrup, R., Hardwick, K.G., Hay, R.T., Kragelund, B.B., Hartmann-Petersen, R., and Gordon, C.** (2014). Dss1 is a 26s proteasome ubiquitin receptor. *Molecular Cell* **56**(3):453–461.

**Pathare, G.R., Nagy, I., Bohn, S., Unverdorben, P., Hubert, A., KÃ¶rner, R., Nickell, S., Lasker, K., Sali, A., Tamura, T., Nishioka, T., FÃ¶rster, F., Baumeister, W., and Bracher, A.** (2012). The proteasomal subunit rpn6 is a molecular clamp

holding the core and regulatory sub-complexes together. *Proceedings of the National Academy of Sciences of the United States of America* **109**(1):149–154.

**Peters, J.M., Harris, J.R., and Kleinschmidt, J.A.** (1991). Ultrastructure of the approximately 26s complex containing the approximately 20s cylinder particle (multi-catalytic proteinase/proteasome). *European Journal of Cell Biology* **56**(2):422–432.

**Rabl, J., Smith, D.M., Yu, Y., Chang, S.C., Goldberg, A.L., and Cheng, Y.** (2008). Mechanism of gate opening in the 20s proteasome by the proteasomal atpases. *Molecular Cell* **30**(3):360–368.

**Ramos, P.C., Häckendorff, J., Johnson, E.S., Varshavsky, A., and Dohmen, R.J.** (1998). Ump1p is required for proper maturation of the 20s proteasome and becomes its substrate upon completion of the assembly. *Cell* **92**(4):489–499.

**Roelofs, J., Park, S., Haas, W., Tian, G., McAllister, F.E., Huo, Y., Lee, B.H., Zhang, F., Shi, Y., Gygi, S.P., and Finley, D.** (2009). Chaperone-mediated pathway of proteasome regulatory particle assembly. *Nature* **459**(7248):861–865.

**Ruschak, A.M., Religa, T.L., Breuer, S., Witt, S., and Kay, L.E.** (2010). The proteasome antechamber maintains substrates in an unfolded state. *Nature* **467**(7317):868–871.

- Russell, J.D., Scalf, M., Book, A.J., Ladror, D.T., Vierstra, R.D., Smith, L.M., and Coon, J.J.** (2013). Characterization and quantification of intact 26s proteasome proteins by real-time measurement of intrinsic fluorescence prior to top-down mass spectrometry. *PLoS One* **8**(3):e58157.
- Sadre-Bazzaz, K., Whitby, F.G., Robinson, H., Formosa, T., and Hill, C.P.** (2010). Structure of a blm10 complex reveals common mechanisms for proteasome binding and gate opening. *Molecular Cell* **37**(5):728–735.
- Saeki, Y., Toh-e, A., Kudo, T., Kawamura, H., and Tanaka, K.** (2009). Multiple proteasome-interacting proteins assist the assembly of the yeast 19s regulatory particle. *Cell* **137**(5):900–913.
- Sakata, E., Bohn, S., Mihalache, O., Kiss, P., Beck, F., Nagy, I., Nickell, S., Tanaka, K., Saeki, Y., Führster, F., and Baumeister, W.** (2012). Localization of the proteasomal ubiquitin receptors rpn10 and rpn13 by electron cryo-microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **109**(5):1479–1484.
- Sakata, E., Stengel, F., Fukunaga, K., Zhou, M., Saeki, Y., Führster, F., Baumeister, W., Tanaka, K., and Robinson, C.V.** (2011). The catalytic activity of ubp6

enhances maturation of the proteasomal regulatory particle. *Molecular Cell* **42**(5):637–649.

**Schliephacke, M., Kremp, A., Schmid, H.P., Kähler, K., and Kull, U.** (1991). Prosomes (proteasomes) of higher plants. *European Journal of Cell Biology* **55**(1):114–121.

**Schmid, H.P., Akhayat, O., Martins, C., Puvion, F., Kähler, K., and Scherrer, K.** (1984). The prosome: a ubiquitous morphologically distinct rnp particle associated with repressed mrnps and containing specific scrna and a characteristic set of proteins. *EMBO Journal* **3**(1):29–34.

**Schmidt, M., Haas, W., Crosas, B., Santamaria, P.G., Gygi, S.P., Walz, T., and Finley, D.** (2005). The heat repeat protein blm10 regulates the yeast proteasome by capping the core particle. *Nature Structural & Molecular Biology* **12**(4):294–303.

**Schreiner, P., Chen, X., Husnjak, K., Randles, L., Zhang, N., Elsasser, S., Finley, D., Dikic, I., Walters, K.J., and Groll, M.** (2008). Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* **453**(7194):548–552.

- Shanklin, J., Jabben, M., and Vierstra, R.D.** (1987). Red light-induced formation of ubiquitin-phytochrome conjugates: identification of possible intermediates of phytochrome degradation. *Proceedings of the National Academy of Sciences of the United States of America* **84**(2):359–363.
- Skoda, B. and Malek, L.** (1992). Dry pea seed proteasome: purification and enzymatic activities. *Plant Physiology* **99**(4):1515–1519.
- Smalle, J.A. and Vierstra, R.D.** (2004). The ubiquitin-26s proteasome proteolytic pathway. *Annual Review of Plant Biology* **55**:555–590.
- Smith, D.M., Kafri, G., Cheng, Y., Ng, D., Walz, T., and Goldberg, A.L.** (2005). Atp binding to pan or the 26s atpsases causes association with the 20s proteasome, gate opening, and translocation of unfolded proteins. *Molecular Cell* **20**(5):687–698.
- Tomko, R.J. and Hochstrasser, M.** (2013). Molecular architecture and assembly of the eukaryotic proteasome. *Annual Review of Biochemistry* **82**:415–445.
- Unverdorben, P., Beck, F., ÅšledÅ°, P., Schweitzer, A., Pfeifer, G., Plitzko, J.M., Baumeister, W., and FÄ¶rster, F.** (2014). Deep classification of a large cryo-em dataset defines the conformational landscape of the 26s proteasome. *Proceedings*

of the National Academy of Sciences of the United States of America **111**(15):5544–5549.

**Van Nocker, S., Deveraux, Q., Rechsteiner, M., and Vierstra, R.D.** (1996). The arabidopsis *mbp1* gene encodes a conserved ubiquitin recognition component of the 26s proteasome. Proceedings of the National Academy of Sciences of the United States of America **93**(2):856–860.

**Verma, R., Aravind, L., Oania, R., McDonald, W.H., Yates, J.R., Koonin, E.V., and Deshaies, R.J.** (2002). Role of the *rpn11* metalloprotease in deubiquitination and degradation by the 26s proteasome. Science **298**(5593):611–615.

**Vierstra, R.D.** (1993). Protein degradation in plants. Annual Review of Plant Physiology and Plant Molecular Biology **44**:385–410.

**Vierstra, R.D.** (2009). The ubiquitin-26s proteasome system at the nexus of plant biology. Nature Reviews Molecular Cell Biology **10**(6):385–397.

**Vierstra, R.D. and Sullivan, M.L.** (1988). Hemin inhibits ubiquitin-dependent proteolysis in both a higher plant and yeast. Biochemistry **27**(9):3290–3295.

**Waxman, L., Fagan, J.M., and Goldberg, A.L.** (1987). Demonstration of two distinct high molecular weight proteases in rabbit reticulocytes, one of which degrades ubiquitin conjugates. Journal of Biological Chemistry **262**(6):2451–2457.

- Weberruss, M.H., Savulescu, A.F., Jando, J., Bissinger, T., Harel, A., Glickman, M.H., and Enenkel, C.** (2013). Blm10 facilitates nuclear import of proteasome core particles. *EMBO Journal* **32**(20):2697–2707.
- Wilkinson, K.D., Urban, M.K., and Haas, A.L.** (1980). Ubiquitin is the atp-dependent proteolysis factor i of rabbit reticulocytes. *Journal of Biological Chemistry* **255**(16):7529–7532.
- Woffenden, B.J., Freeman, T.B., and Beers, E.P.** (1998). Proteasome inhibitors prevent tracheary element differentiation in zinnia mesophyll cell cultures. *Plant Physiology* **118**(2):419–429.
- Worden, E.J., Padovani, C., and Martin, A.** (2014). Structure of the rpn11-rpn8 dimer reveals mechanisms of substrate deubiquitination during proteasomal degradation. *Nature Structural & Molecular Biology* **21**(3):220–227.
- Xie, Y. and Varshavsky, A.** (2000). Physical association of ubiquitin ligases and the 26s proteasome. *Proceedings of the National Academy of Sciences of the United States of America* **97**(6):2497–2502.
- Yanagawa, Y., Ohhashi, A., Murakami, Y., Saeki, Y., Yokosawa, H., Tanaka, K., Hashimoto, J., Sato, T., and Nakagawa, H.** (1999). Purification and characteriza-

tion of the 26s proteasome from cultured rice (*oryza sativa*) cells. *Plant Science* **149**(1):33–41.

**Yang, P., Fu, H., Walker, J.M., Papa, C.M., Smalle, J., Ju, Y.M., and Vierstra, R.D.** (2004). Purification of the arabidopsis 26s proteasome: biochemical and molecular analyses revealed the presence of multiple isoforms. *Journal of Biological Chemistry* **279**(8):6401–6413.

**Yoshimura, T., Kameyama, K., Takagi, T., Ikai, A., Tokunaga, F., Koide, T., Tanahashi, N., Tamura, T., Cejka, Z., Baumeister, W., Tanaka, K., and Ichihara, A.** (1993). Molecular characterization of the 26s proteasome complex from rat liver. *Journal of Structural Biology* **111**(3):200–211.

## Chapter 2

### **MORPHEUS SPECTRAL COUNTER: A COMPUTATIONAL TOOL**

### **FOR LABEL-FREE QUANTITATIVE MASS SPECTROMETRY**

### **USING THE MORPHEUS SEARCH ENGINE**

#### **2.1. Summary**

Label-free quantitative MS based on the Normalized Spectral Abundance Factor (NSAF) has emerged as a straightforward and robust method to determine the relative abundance of individual proteins within complex mixtures. Here, we present Morpheus Spectral Counter (MSpC) as the first computational tool that directly calculates NSAF values from output obtained from Morpheus, a fast, open-source, peptide-MS/MS matching engine compatible with high-resolution accurate-mass instruments. NSAF has distinct advantages over other MS-based quantification methods, including a higher dynamic range as compared to isobaric tags, no requirement to align and re-extract MS1 peaks, and increased speed. MSpC features an easy to use graphic user interface that additionally calculates both distributed and unique NSAF values to permit analyses of both protein families and isoforms/proteoforms. MSpC determinations of protein concentration were linear over

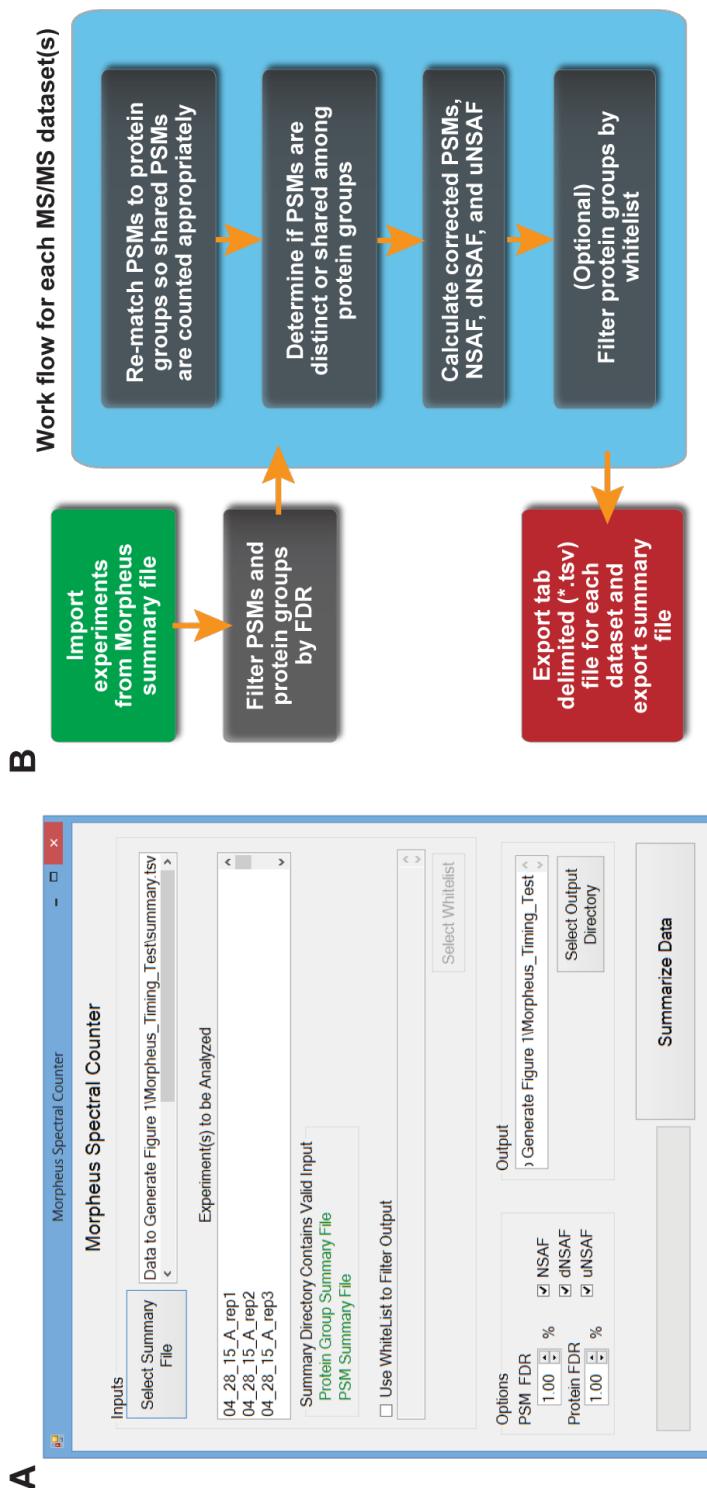
several orders of magnitude based on the analysis of several high-mass accuracy datasets either obtained from PRIDE or generated with total cell extracts spiked with purified *Arabidopsis* 20S proteasomes. The MSpC software was developed in C# and is open sourced under a permissive license with the code made available at [http://dcgempeline.github.io/Morpheus\\_SpC/](http://dcgempeline.github.io/Morpheus_SpC/).

## 2.2. Main Text

Quantification of individual polypeptides within complex mixtures by MS is an extremely useful tool to understand proteomic changes in organisms during growth and development, and after environmental perturbation (Wong and Cagney, 2010). While a number of MS/MS strategies have been developed to measure protein abundance, including Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC), labeling with isobaric tags, and Absolute Quantification of proteins (AQUA) (Gerber et al., 2003; Ong et al., 2002; Ross et al., 2004; Thompson et al., 2003), label-free quantification (LFQ) have become increasingly popular given their simplicity and low cost (Wong and Cagney, 2010; Zhang et al., 2006). One LFQ strategy infers abundance from the number of observed peptide spectra matches (PSMs). For these PSM-based approaches, changes in protein abundance can be generated artificially when total PSMs differ among samples and because longer proteins

tend to produce more raw counts. For these reasons normalizing for both protein length and total PSMs is paramount. While this adjustment can be made in a number of ways; one of the most straight forward methods is to use Normalized Spectral Abundance Factor (NSAF), a length- and count-normalized measure for each protein (Zybailov et al., 2006). Further improvements to the NSAF algorithm have been made by accounting for shared peptides in distributed NSAF (dNSAF), which distributes common PSMs among a family of isoforms/proteoforms based on the number of distinct PSMs observed for each isoform/proteoform, and unique NSAF (uNSAF), which ignores shared PSMs and only assigns distinct PSMs to each specific isoform/proteoform (Zhang et al., 2010).

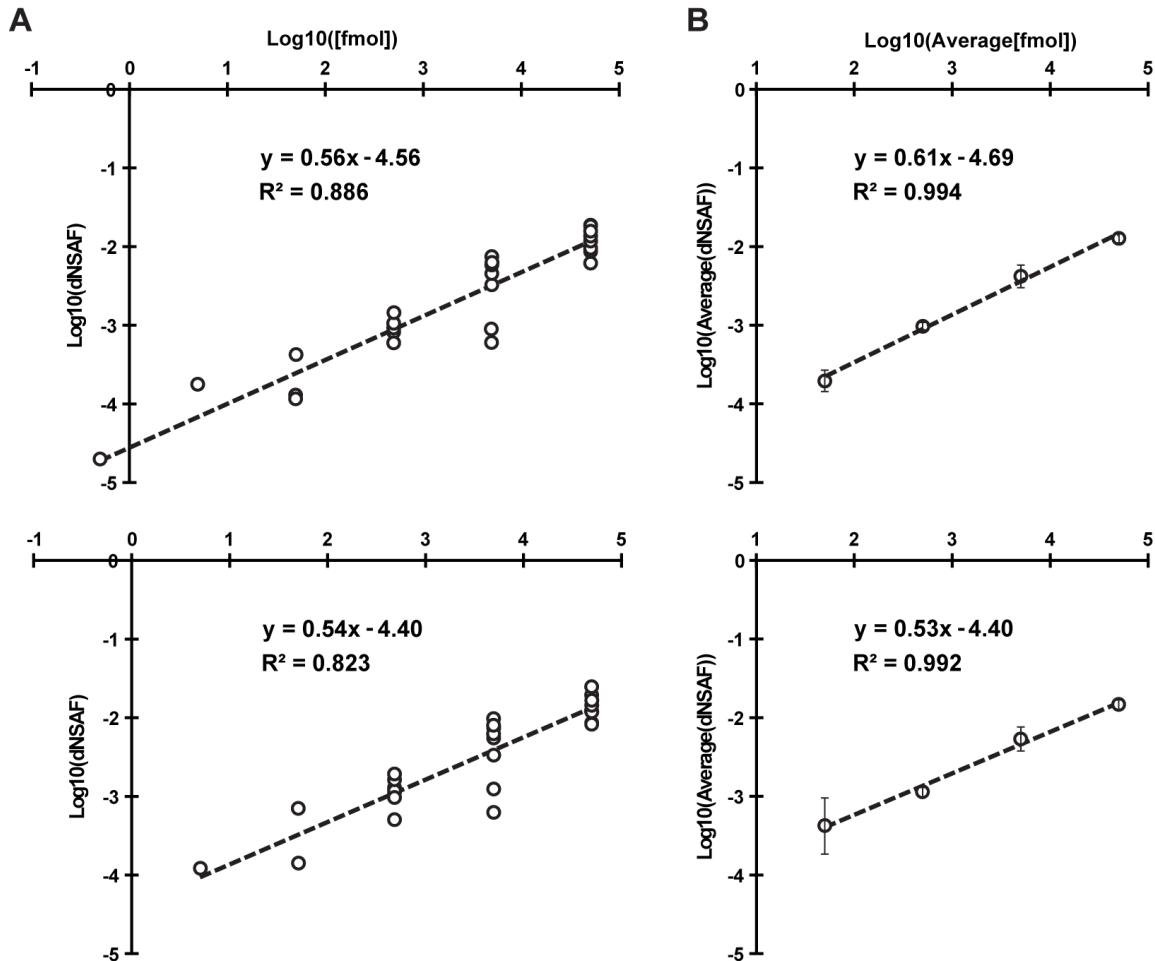
The Morpheus MS search engine was recently designed for high-resolution, accurate-mass data obtained from Orbitrap-based instruments to provide faster matching of spectra to peptides (Wenger and Coon, 2013). Unfortunately, no downstream automated tools are available to facilitate LFQ analysis, which can be quite challenging, if not impossible, to complete manually when accounting for shared peptides. To overcome this bottleneck, we developed Morpheus Spectral Counter (MSpC) as the first LFQ computational tool that integrates directly with Morpheus to calculate NSAF, dNSAF, uNSAF, and corrected PSM (Fermin et al., 2011) values in complex protein samples. MSpC is fully automated, and only requires a



**Figure 2.1: MSpC Graphic User Interface (GUI) and software flow chart.** (A) Screenshot of the GUI. The input requires the user to select a Morpheus search summary file containing experiments to be analyzed. The user can optionally select a whitelist to filter output, and select an output directory. Additional options can set peptide and protein FDR cutoffs, and method of quantification for output, including Normalized Spectral Abundance Factor (NSAF), distributed NSAF (dNSAF), and unique NSAF (uNSAF). A progress bar highlights completion of the analysis. (B) Data analysis flow chart. Experiments and groups of experiments to be analyzed are imported through the Morpheus summary file. PSMs and protein groups are filtered at the specified FDR cutoff with a default of 1%.

Morpheus search summary file (summary.tsv) as input. The user interface (see Figure 2.1 (A)) allows one to select the summary file and displays the raw MS/MS files that will be analyzed by MSpC. Due to shared peptides being attributed to only one instance of a protein group in Morpheus's PSM file, PSMs are re-matched to all possible protein groups. PSMs are then cataloged as shared or as unique (distinctly matching one protein group) to generate NSAF, dNSAF, and uNSAF outputs. Finally, the output can be filtered for proteins of interest by specifying a comma delimited file containing unique identifiers and descriptions. Some important features of MSpC are its ability to handle fractionation experiments as input, and the ability to whitelist proteins of interest in the output by specifying a csv file (see Tutorial 2.5). Options exist to specify global PSM and protein group FDR rates (thus avoiding increased FDRs when one analyzes many experiments at once), to output NSAF, dNSAF, and uNSAF values, to require a minimum number of unique peptides to quantify a protein, and to specify an output directory. A progress bar indicates completion of the analysis by MSpC. To validate the accuracy of MSpC, we analyzed two MS/MS datasets available in PRIDE that were previously generated by high-energy collision-induced dissociation using Thermo Q-Exactive Orbitrap instruments.

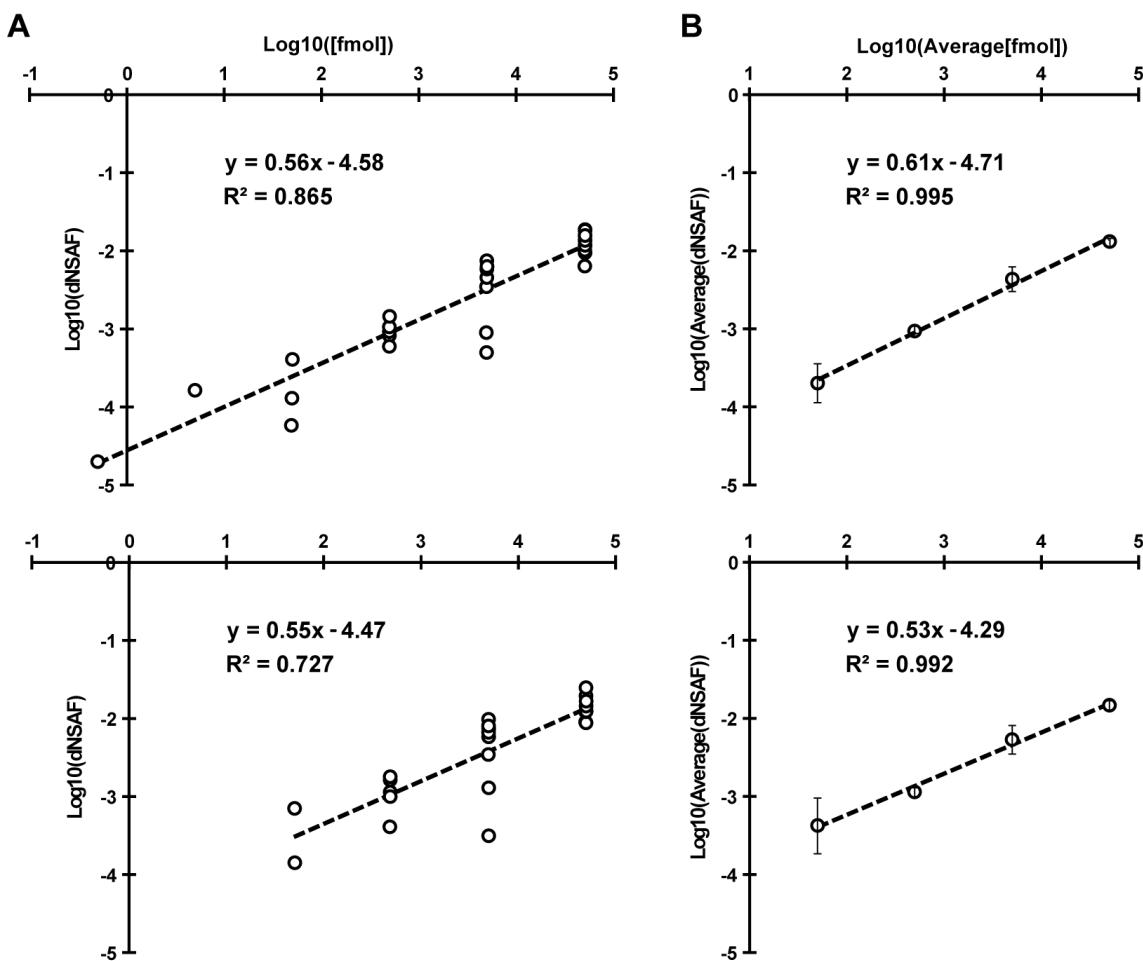
Here, Xenopus egg (see top, Figure 2.2) and embryo (bottom, Figure 2.2) extracts



**Figure 2.2: Confirmation of MSpC accuracy by analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2).** The array of UPS2 standards were spiked into *Xenopus laevis* egg (**Top**) and embryo (**Bottom**) extracts at a range of concentrations. Following MS/MS analysis, dNSAF values for each protein were determined by Morpheus and MSpC. (**A**) A log-log plot of dNSAF versus concentration for each UPS2 protein detected across each fmol range. (**B**) A log-log plot of average dNSAF vs average concentration of each group of UPS2 proteins at each fmol range: (50, 500, 5000, and 50,000 fmol).

were spiked at a 4:1 ratio with the Universal Proteome Standard 2 (UPS2), a mix of 48 purified proteins at defined molar ratios of 0.5, 5, 50, 500, 5000, and 50,000, with each ratio containing a different set of 8 of the 48 proteins. As shown in Figure 2.2 A, when the Morpheus/MSpC pipeline was used to calculate the average dNSAF value for each UPS2 protein, requiring only a single unique peptide to quantify, strong linear correlations ( $R^2 = 0.886$  and  $0.823$ ) were obtained across a 1,000 fold change in abundance (50 fmol to 50,000 fmol). In fact, the  $R^2$  values were similar to those obtained by others with PSM-based LFQ methods (Cox et al., 2014; Tu et al., 2014). This linear correlation was further strengthened when the dNSAF values were averaged for all UPS2 proteins within each of the concentration groups, with  $R^2$  values of 0.994 and 0.992 for the egg and embryo datasets, respectively (see Figure 2.2 B). Notably, the slope of the concentration series was significantly less than unity, showing that NSAF measurements are not appropriate for absolute quantification, which was expected given that NSAF is a relative value.

We also reprocessed the UPS2 dataset using the option of requiring a minimum of two unique peptides for quantification, which should improve stringency. This option provided only a minor improvement in overall linearity for the average UPS2 dNSAF values, but decreased linearity when each UPS2 protein was considered individually and removed some UPS2 proteins at low concentrations (compare

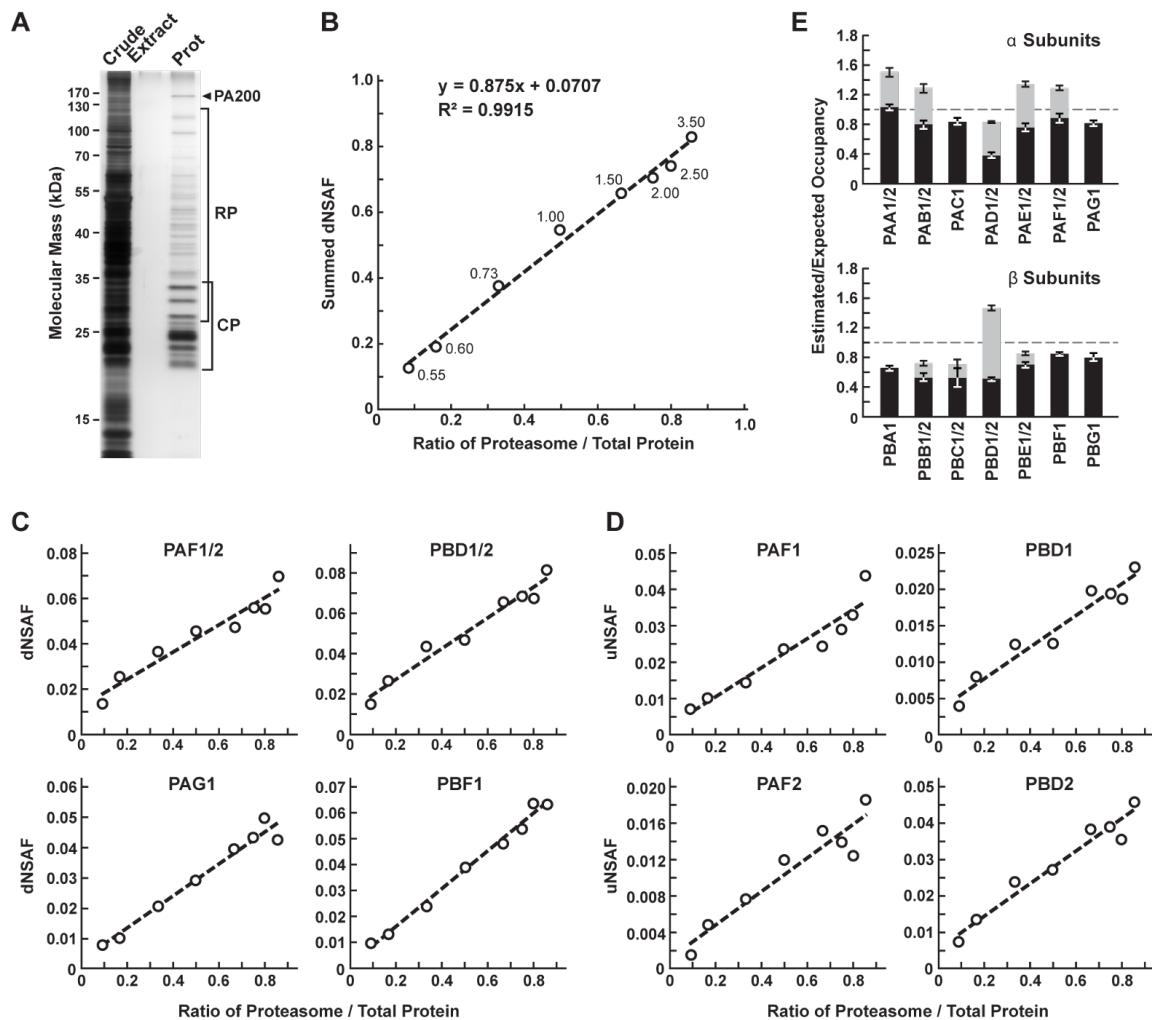


**Figure 2.3: Re- Analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2).** The array of UPS2 standards were spiked into Xenopus laevis egg (**Top**) and embryo (**Bottom**) extracts at a range of concentrations. Following MS/MS analysis, dNSAF values for each protein were determined by Morpheus and MSpC with a change from Figure 2.2 in that two unique peptides were required to quantify a protein. **(A)** A log-log plot of dNSAF versus concentration for each UPS2 protein detected across each fmol range. **(B)** A log-log plot of average dNSAF vs average concentration of each group of UPS2 proteins at each fmol range: (50, 500, 5000, and 50,000 fmol).

Figure 2.3 A to Figure 2.2 A). Consequently, caution should be exercised when selecting this option even though it might provide a slight improvement in stringency (see Discussion on Requiring Two Peptide 2.4). To demonstrate the utility and accuracy of MSpC as applied to our work, we analyzed 20S proteasomes isolated from *Arabidopsis thaliana*. This particle contains multiple subunits assembled in stoichiometric amounts, with many subunits encoded by two paralogous genes of sufficient amino acid identity (typically >90% (Yang et al., 2004)) such that discrimination between paralogs can be challenging using LFQ approaches (Book et al., 2010). To simulate changes in 20S proteasome abundance, we added varying amounts of trypsinized proteasomes (0.05 µg to 3 µg) to a fixed amount of trypsinized *Escherichia coli* lysate (0.5 µg) to generate proteasome/lysate ratios of ~0.091, 0.167, 0.333, 0.500, 0.667, 0.750 0.800, 0.857. The digests were then subjected to MS/MS and the dNSAF value for each subunit along with the uNSAF value for individual isoforms were calculated by the Morpheus/MSpC pipeline (see Methods 2.3). The data from this experiment are deposited in PRIDE with ID PXD003002. As shown in Figure 2.4, MSpC provided an excellent determination for the overall abundance of 20S proteasomes within a complex mixture, along with a good reflection of the abundance of individual subunits and their isoforms.

When the dNSAF values for all subunits for the *Arabidopsis* 20S proteasome

[Figure 2.4 caption follows on next page]



including their isoforms (representing 14 distinct subunits, 10 of which exist as isoform pairs) were summed, a very close approximation of the dNSAF/actual abundance was obtained (slope=0.875) with a very strong linear correlation ( $R^2 = 0.99$ ) over a ~10-fold range in protein abundance. When each 20S proteasome subunit was analyzed individually, a strong linear response was also obtained ( $R^2 > 0.90$ ) for a majority of subunits (Figure 2.4 C and Table 2.1).

For example, reasonably accurate concentration plots were obtained for the

---

**Figure 2.4 (preceding page): Confirmation of MSpC accuracy by analysis of MS/MS datasets generated with affinity purified *Arabidopsis* 20S proteasomes spiked into a total cell lysate from *E. coli*. Following MS/MS analysis, the dNSAF and uNSAF values for each subunit/isoform were determined by Morpheus and MSpC.** (A) A silver-stained SDS-PAGE gel of 20S proteasome samples affinity purified from 10-d-old *Arabidopsis* seedlings. The crude seedling extract(CE), sample buffer (SB), and affinity-purified 20S proteasome samples (Prot) are shown. (B) Quantification of trypsinized 20S proteasomes when mixed at varying ratios with trypsinized total protein lysates from *E. coli*. The spiked samples were subjected to MS/MS followed by data analysis with the Morpheus and MSpC. dNSAF values for each proteasome subunit were averaged across three technical replicates, then summed to obtain an estimate of abundance for the 20S proteasome, and plotted against their known ratios. The total protein load is listed at each point in ug. (C and D) dNSAF and uNSAF values determined from the data in panel B for individual subunits (C) and their isoforms (D) for several subunits of the 20S proteasome. (E) Quantification accuracy of the Morpheus/MSpC pipeline for determining the amount of each  $\alpha$  and  $\beta$  subunit of the 20S proteasome. Single subunit isoforms are in black, whereas subunits having two isoforms are shown in black and grey to reflect the contributions of isoforms 1 and 2 respectively. Each bar represents the average of eight technical replicates ( $\pm$  SE). The dashed line represents the expected value of one assuming an equal stoichiometry of each subunit within the particle.

PAF ( $\alpha 6$ ) and PBD ( $\beta 4$ ) subunits that are encoded by the PAF1/2 and PBF1/2 gene pairs, and for the PAG ( $\alpha 7$ ) and PBF ( $\beta 6$ ) subunits that are encoded by single PAG1 and PBF1 genes ( $R^2$  from 0.94 to 0.99). Even when we calculated uNSAF values for individual isoforms added to the *E. coli* lysate, strong linear responses were obtained (e.g., the PAF1/PAF2 and PBD1/PBD2 pairs) with robust correlations ( $R^2$  from 0.89 to 0.95) (Figure fig:proteasomespike D). Taken together, MSpC worked well for relative LFQ analysis of a multi-subunit complex and its individual subunits and isoforms within a complex proteomic mixture.

The Morpheus/MSpC pipeline also allowed us to calculate the respective incorporation of each paralog in the complex (see Isoform Incorporation Methods 2.3.4). As shown in Figure 2.4E, these estimated/expected occupancies were close to unity for most subunits within both the  $\alpha$  and  $\beta$  rings of the 20S proteasome. The only strong deviation was for PBD1/2 ( $\beta 4$ ), which had a greater dNSAF value relative to other  $\beta$  subunits across the experiments analyzed (see Table 2.1). The calculations for uNSAF values also estimated the relative proportion of each isoform within the complex for those subunits expressed from paralogous genes. The data obtained are similar to prior studies of the complex involving quantitative top-down proteomic analysis of purified proteasome samples using ultra violet-intrinsic fluorescence to quantify tyrosine-containing subunits (Russell et al., 2013). However, our MSpC

analysis provided a more complete picture as several subunit isoforms were difficult to quantify by fluorescence either because they lacked tryosine, or because their fluorescence peaks overlapped with those of other subunits/isoforms. Notably , the protein isoform ratios measured here agree well with the expression ratios for the paralogous genes (Book et al., 2010), suggesting that the protein isoform abundance generally reflects the relative transcriptional activity of the gene pair. We consistently estimated slightly more  $\alpha$  ring subunits (PAA-PAG) versus  $\beta$  ring subunits (PBA-PBG) in the final MSpC calculations (Figure 2.4 E). This deviation could represent enhanced detection of  $\alpha$  ring versus  $\beta$  ring subunits, or more likely that purification via the tagged  $\alpha$  ring subunit PAG1 also isolated assembly intermediates comprised of only  $\alpha$  ring subunits.

We compared the Morpheus and MSpC pipeline to the next most comparable open source, spectral-count-based LFQ pipeline, The Trans Proteomic Pipeline (TPP) (Deutsch et al., 2010) and ABACUS (Fermin et al., 2011) using our datasets generated with the 20S proteasome/*E. coli* lysate mixture (see Tables 2.1 and 2.2). Morpheus/MSpC slightly outperformed TPP/ABACUS by having a greater overall accuracy (average linearity of 0.88 compared to 0.84), and by having more subunits showing an  $R^2$  linear correlation greater than 0.9 (14/23 subunits for MSpC versus and 11/23 for ABACUS).

**Table 2.1: Table of dNSAF values for each 20S proteasome subunit generated by analyzing the proteasome spike in experiments with the Morpheus and MSpC pipeline.** The top half of the table lists  $\alpha$ 1-7 (PAA-PAG) where 1 or 2 represent different isoforms) subunits, while the bottom half of the table lists  $\beta$ 1-7 subunits (PBA-PBG where 1 or 2 represent different isoforms)

| Ratio | 0.091  | 0.167  | 0.333  | 0.500  | 0.667  | 0.750  | 0.800  | 0.857  |         | Pearsons |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|---------|----------|
| PAA1  | 0.0102 | 0.0180 | 0.0275 | 0.0392 | 0.0505 | 0.0489 | 0.0467 | 0.0530 |         | 0.952    |
| PAA2  | 0.0063 | 0.0072 | 0.0118 | 0.0121 | 0.0139 | 0.0218 | 0.0351 | 0.0195 |         | 0.652    |
| PAB1  | 0.0091 | 0.0117 | 0.0209 | 0.0391 | 0.0377 | 0.0377 | 0.0268 | 0.0397 |         | 0.721    |
| PAB2  | 0.0017 | 0.0062 | 0.0147 | 0.0223 | 0.0295 | 0.0277 | 0.0303 | 0.0250 |         | 0.894    |
| PAC1  | 0.0076 | 0.0092 | 0.0210 | 0.0339 | 0.0373 | 0.0395 | 0.0546 | 0.0512 |         | 0.955    |
| PAD1  | 0.0056 | 0.0070 | 0.0103 | 0.0151 | 0.0160 | 0.0147 | 0.0149 | 0.0147 |         | 0.826    |
| PAD2  | 0.0039 | 0.0068 | 0.0141 | 0.0162 | 0.0200 | 0.0225 | 0.0255 | 0.0228 |         | 0.959    |
| PAE1  | 0.0046 | 0.0095 | 0.0208 | 0.0352 | 0.0395 | 0.0356 | 0.0389 | 0.0490 |         | 0.928    |
| PAE2  | 0.0063 | 0.0082 | 0.0184 | 0.0240 | 0.0269 | 0.0282 | 0.0266 | 0.0293 |         | 0.925    |
| PAF1  | 0.0106 | 0.0173 | 0.0239 | 0.0303 | 0.0292 | 0.0379 | 0.0406 | 0.0492 |         | 0.920    |
| PAF2  | 0.0032 | 0.0083 | 0.0128 | 0.0155 | 0.0182 | 0.0184 | 0.0152 | 0.0209 |         | 0.847    |
| PAG1  | 0.0080 | 0.0100 | 0.0208 | 0.0291 | 0.0394 | 0.0431 | 0.0498 | 0.0423 |         | 0.966    |
| Ratio | 0.091  | 0.167  | 0.333  | 0.500  | 0.667  | 0.750  | 0.800  | 0.857  |         | Pearsons |
| PBA1  | 0.0076 | 0.0104 | 0.0170 | 0.0207 | 0.0246 | 0.0288 | 0.0337 | 0.0449 |         | 0.898    |
| PBB1  | 0.0066 | 0.0074 | 0.0191 | 0.0276 | 0.0223 | 0.0206 | 0.0172 | 0.0227 |         | 0.498    |
| PBB2  | 0.0009 | 0.0016 | 0.0034 | 0.0035 | 0.0110 | 0.0109 | 0.0160 | 0.0161 |         | 0.891    |
| PBC1  | 0.0000 | 0.0000 | 0.0162 | 0.0309 | 0.0373 | 0.0344 | 0.0295 | 0.0388 |         | 0.877    |
| PBC2  | 0.0000 | 0.0000 | 0.0000 | 0.0096 | 0.0144 | 0.0147 | 0.0146 | 0.0190 |         | 0.931    |
| PBD1  | 0.0053 | 0.0097 | 0.0150 | 0.0147 | 0.0221 | 0.0228 | 0.0231 | 0.0271 |         | 0.955    |
| PBD2  | 0.0094 | 0.0162 | 0.0286 | 0.0318 | 0.0426 | 0.0457 | 0.0441 | 0.0538 |         | 0.969    |
| PBE1  | 0.0070 | 0.0097 | 0.0156 | 0.0224 | 0.0323 | 0.0373 | 0.0334 | 0.0519 |         | 0.914    |
| PBE2  | 0.0007 | 0.0003 | 0.0036 | 0.0069 | 0.0080 | 0.0105 | 0.0106 | 0.0133 |         | 0.973    |
| PBF1  | 0.0093 | 0.0122 | 0.0205 | 0.0310 | 0.0368 | 0.0412 | 0.0473 | 0.0459 |         | 0.991    |
| PBG1  | 0.0085 | 0.0165 | 0.0203 | 0.0323 | 0.0331 | 0.0345 | 0.0334 | 0.0442 |         | 0.912    |
|       |        |        |        |        |        |        |        |        | Average | 0.885    |

In addition to this modest improvement, we note that the Morpheus/MSpC pipeline required significantly less intermediary steps, thus accelerating the data analysis. Some of the additional steps in TPP/ABACUS could be automated from the command-line, but it would likely be a challenge for the average user. Importantly, we found that the Morpheus/MSpC pipeline was faster. Timing tests using the proteasome/*E. coli* spike data generated here showed that the Morpheus/MSpC

**Table 2.2: Table of adj\_NSAF values for each 20S proteasome subunit (equivalent to dNSAF) generated by analyzing the proteasome spike in experiments with the TPP and ABACUS pipeline.** The top half of the table lists  $\alpha$ 1-7 (PAA-PAG) where 1 or 2 represent different isoforms) subunits, while the bottom half of the table lists  $\beta$ 1-7 subunits (PBA-PBG where 1 or 2 represent different isoforms)

| Ratio | 0.091  | 0.167  | 0.333  | 0.500  | 0.667  | 0.750  | 0.800  | 0.857  |         | Pearsons |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|---------|----------|
| PAA1  | 125.16 | 166.81 | 245.61 | 355.60 | 491.76 | 494.18 | 504.75 | 554.97 |         | 0.988    |
| PAA2  | 57.52  | 70.52  | 114.60 | 126.35 | 132.06 | 162.25 | 358.13 | 154.62 |         | 0.493    |
| PAB1  | 96.13  | 134.98 | 207.90 | 328.16 | 362.21 | 416.46 | 303.60 | 397.39 |         | 0.865    |
| PAB2  | 29.67  | 79.42  | 153.34 | 282.91 | 365.22 | 299.47 | 310.19 | 299.93 |         | 0.849    |
| PAC1  | 95.27  | 112.45 | 214.24 | 376.66 | 389.05 | 361.39 | 512.73 | 488.62 |         | 0.925    |
| PAD1  | 62.58  | 76.16  | 106.03 | 196.03 | 179.37 | 182.99 | 164.04 | 188.71 |         | 0.794    |
| PAD2  | 39.88  | 68.33  | 145.34 | 178.43 | 190.84 | 246.20 | 243.28 | 240.01 |         | 0.954    |
| PAE1  | 50.41  | 97.53  | 193.18 | 368.87 | 464.79 | 397.27 | 468.64 | 534.16 |         | 0.951    |
| PAE2  | 71.38  | 80.07  | 184.85 | 255.98 | 299.41 | 290.16 | 304.77 | 348.32 |         | 0.955    |
| PAF1  | 132.67 | 179.60 | 207.68 | 253.51 | 253.04 | 327.11 | 334.73 | 460.76 |         | 0.835    |
| PAF2  | 0.00   | 0.00   | 59.62  | 70.93  | 111.83 | 80.29  | 63.08  | 65.86  |         | 0.606    |
| PAG1  | 108.09 | 142.82 | 271.18 | 389.94 | 495.94 | 530.90 | 632.72 | 540.86 |         | 0.966    |
| <hr/> |        |        |        |        |        |        |        |        |         |          |
| Ratio | 0.091  | 0.167  | 0.333  | 0.500  | 0.667  | 0.750  | 0.800  | 0.857  |         | Pearsons |
| PBA1  | 85.99  | 113.01 | 208.38 | 233.44 | 273.50 | 327.23 | 417.12 | 560.62 |         | 0.851    |
| PBB1  | 47.77  | 80.66  | 195.35 | 324.53 | 303.26 | 220.36 | 200.51 | 216.77 |         | 0.440    |
| PBB2  | 17.11  | 4.54   | 0.00   | 0.00   | 21.05  | 102.51 | 144.30 | 187.86 |         | 0.618    |
| PBC1  | 27.19  | 40.65  | 165.46 | 345.14 | 364.19 | 386.59 | 338.47 | 373.16 |         | 0.885    |
| PBC2  | 10.64  | 40.65  | 0.00   | 55.58  | 145.43 | 158.90 | 159.74 | 228.05 |         | 0.845    |
| PBD1  | 50.14  | 94.46  | 168.58 | 167.50 | 244.25 | 248.17 | 252.65 | 319.61 |         | 0.945    |
| PBD2  | 81.45  | 149.28 | 256.88 | 319.89 | 390.60 | 407.02 | 380.30 | 490.17 |         | 0.950    |
| PBE1  | 81.01  | 107.76 | 196.21 | 302.90 | 365.03 | 404.25 | 378.67 | 441.03 |         | 0.981    |
| PBE2  | 0.00   | 1.52   | 5.59   | 8.05   | 40.06  | 56.56  | 59.39  | 58.56  |         | 0.887    |
| PBF1  | 71.53  | 112.26 | 190.45 | 226.89 | 259.90 | 366.60 | 431.42 | 420.23 |         | 0.943    |
| PBG1  | 104.93 | 174.55 | 232.34 | 392.37 | 425.10 | 397.85 | 396.42 | 463.13 |         | 0.906    |
|       |        |        |        |        |        |        |        |        | Average | 0.845    |

pipeline was 1.9-fold faster than the TPP/ABACUS pipeline (Figure 2.5). Such an improvement was expected given that Morpheus completes its searches on average 1.3 to 4.6 times faster than most other search engines available (Wenger and Coon, 2013). Given its simplicity of use, speed, and open source nature, MSpC combined with Morpheus is clearly advantageous over other PSM-based LFQ approaches currently available. Moreover, by being open source, MSpC should allow others to

extend its utility and to serve as a platform for integrating additional open source LFQ approaches into the Morpheus pipeline.

## 2.3. Methods

### 2.3.1. Sample Preparation

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. 20S proteasomes were obtained as described previously (Book et al., 2010) from *Arabidopsis thaliana* Col-0 ecotype seedlings in which the 20S proteasome subunit PAG1 ( $\alpha 7$ ) was genetically replaced with a FLAG-tagged variant, with the minor modification of switching to a more stable HEPES buffer during purification. The FLAG peptide used for elution was removed by filtering through an Amicon Ultra 4 10K filter with the elution buffer also exchanged into 8 M urea. Total protein was quantified by the bicinchoninic acid protein assay (Thermo Scientific) using bovine serum albumin as the standard. Approximately 70  $\mu$ g of proteasomes were digested overnight at 37°C using a 1:30 trypsin/sample ratio. Peptides were acidified to a final concentration of 1% TFA, desalted on a Waters C18 Sep-Pak containing 50 mg sorbent material, and lyophilized. Total *E. coli* lysates were obtained from Bio-Rad (Cat. 163-2110) with 200  $\mu$ g digested as above. Both proteasome and *E. coli* peptides were dissolved in 5% acetonitrile, 95% water, and 0.1 % formic acid. Each MS analysis, performed

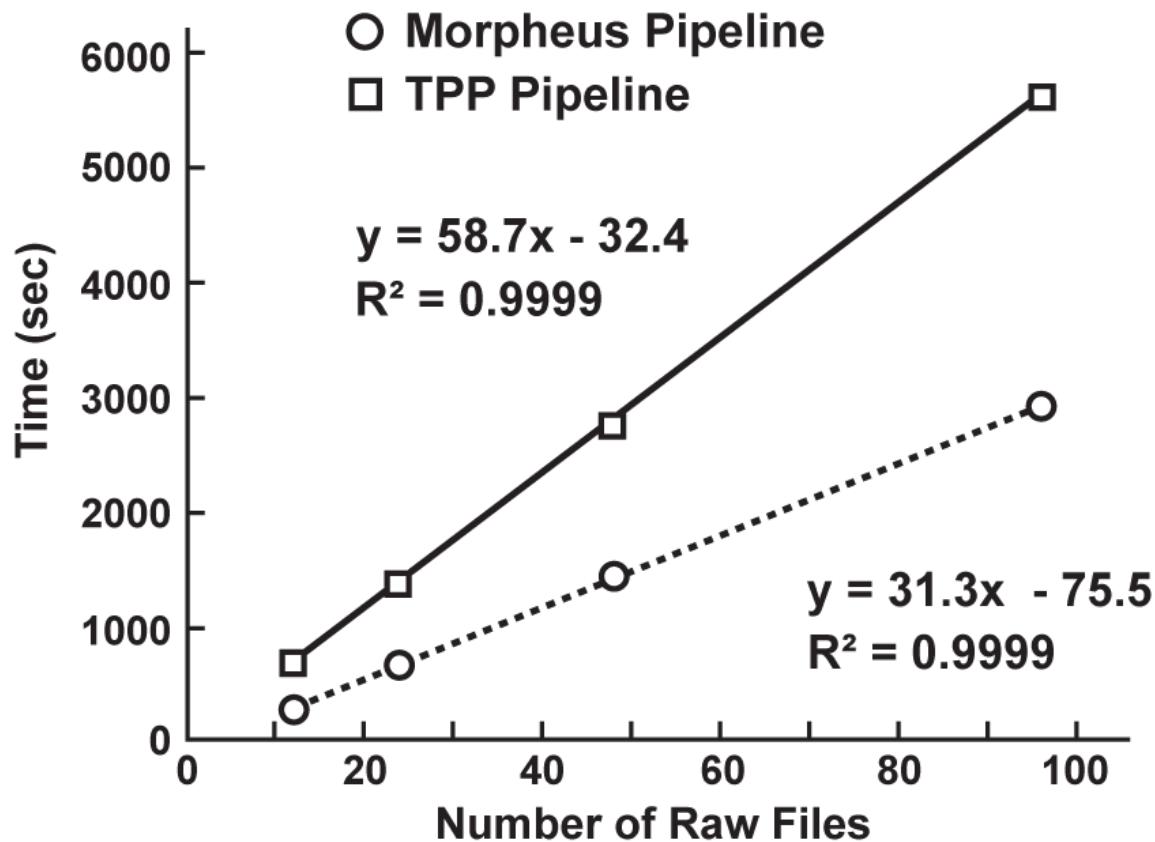


Figure 2.5: MSpC combined with Morpheus works faster than TPP combined with ABACUS. Speed comparisons were performed for 12, 24, 48, and 96 raw MS/MS files generated with the 20S proteasome/*E. coli* lysate samples analyzed in 2.4. On average, Morpheus/MSpC finished the calculations 1.9 times faster than TPP/ABACUS over a ~10-fold range of dataset size.

in triplicae, used 5  $\mu$ L volumes prepared with 3, 2, 1.5, 1, 0.5, 0.25, 0.1, or 0.05  $\mu$ g of digested proteasomes mixed with 0.5  $\mu$ g of digested *E. coli* proteins. This mixtures reflected proteasome/*E. coli* ratios of ~0.091, 0.167, 0.333, 0.500, 0.667, 0.750 0.800, and 0.857, respectively.

### **2.3.2. Liquid Chromatography and High-Resolution Mass Spectrometry**

Samples were analyzed by ultra-high performance liquid chromatography (UPLC) (nanoAcuity, Waters Corporation) connected online to an electrospray ionization LTQ Velos Orbitrap mass spectrometer (Thermo Scientific). Separation employed a 100 x 365  $\mu$ m fused silica capillary micro-column packed with 20 cm of 1.7  $\mu$ m-diameter, 130- $\text{\AA}$  pore size, C18 beads (Waters BEH), with an emitter tip pulled to approximately 1  $\mu$ m using a laser puller (Sutter Instruments). Peptides were loaded at a flow-rate of 400 nL/min for 30 min and then eluted over 120 min at a flow-rate of 300 nL/min with a 2% to 30% acetonitrile gradient in 0.1% formic acid. Full-mass scans were performed in the FT Orbitrap with a mass range of 300-1500 m/z at a resolution of 60,000, followed by ten MS/MS high energy C-trap dissociation scans of the ten highest intensity parent ions at 42% normalized collision energy and 7,500 resolution, with a mass range starting at 100 m/z. Dynamic exclusion was enabled with a repeat count of two over the duration of 30 sec and an exclusion

window of 120 sec.

### 2.3.3. Data Processing

Protein identifications were determined using the Morpheus search engine (Wenger and Coon, 2013). Raw data was searched with the Thermo module of Morpheus revision 151 downloaded and compiled from source code available at <http://sourceforge.net/projects/morpheus-ms/> using Microsoft Visual Studio 2013 professional edition. The following parameters were used to search all databases: unknown precursor charge states - +2, +3, +4; maximum number of MS/MS peaks = 400; assign charge states - enabled; de-isotope - disabled; generate target decoy database on the fly; protease trypsin (no proline rule); maximum missed cleavages = 2; initiator methionine behavior - variable; fixed modification of carbamidomethylation of cysteine; variable modification of oxidation of methionine; maximum variable modification isoforms per peptide = 1024; precursor mass tolerance = ± 2.1 Da monoisotopic (recommended parameters to account for neutral loss); precursor monoisotopic peak correction - disabled; product mass tolerance = ± 0.025 Da monoisotopic; consider modified forms as unique peptides - false; maximum threads = 8; minimize memory usage - false. MSpC quantification of Universal Proteome Standard 2 (UPS2) protein sequences exploited the MS/MS analysis of

individual USP2 proteins (Sigma-Aldrich) mixed at various concentrations with egg or embryo extracts from *Xenopus laevis* available in the PRoteomics IDEntifications (PRIDE) repository (Vizcaino et al., 2013) using identifier - PXD000902 and the available proteomics database - pita\_v1.71.protein.name.fa. For our analysis, both the raw MS/MS data and resulting FASTA files for the USP2 and egg and embryo proteomes were obtained from PRIDE. The database used for searching the MS/MS data of *Arabidopsis* proteasomes spiked into total *E. coli* peptides was generated by combining Uniprot K12 *E. coli* reference proteome UP000000625 with a common contaminant database, and then mixing the merged dataset with FASTA sequences for all proteoforms of all known proteasome subunits and associated proteins (Book et al., 2010) obtained from the TAIR10\_pep\_20101214 FASTA database available within The *Arabidopsis* Information Resource (TAIR) version 10. All FASTA files are available for download in the Supporting Information. The datasets were analyzed by MSpC with a 1% PSM false discovery rate (FDR) and a 1% protein group FDR to determine NSAF, dNSAF, and uNSAF values for each protein group. Two separate analyses were performed in which one unique, or alternatively two unique peptides were required to quantify a protein group. NSAF values were calculated according to formulas 1a, 2a, and 3a from Figure 1 of Zhang et al. (Zhang et al., 2010).

#### **2.3.4. Isoform Incorporation Rates**

For the individual subunit analysis of the 20S proteasome the isoform incorporation rates were treated as follows. Given that each proteasome subunit should be incorporated at equal stoichiometry within the 20S particle, we then tested whether the Morpheus/MSpC pipeline could calculate the relative abundance of each subunit and the distribution of isoforms. Here, we divided the dNSAF values for each subunit/isoform by the total number of dNSAF values for the entire complex across all eight total proteasome/*E. coli* lysate ratios tested. This averaged value provided a concentration-independent ratio for the incorporation of each subunit/isoform. We then normalized these values based on a 1/14 stoichiometry of each subunit within the complex to calculate the estimated occupancy versus the expected occupancy of each subunit.

#### **2.3.5. Speed and Accuracy Comparisons of Morpheus/MSpC to the TPP/ABA-CUS Pipeline**

The speed and accuracy of MSpC combined with Morpheus was compared to the next most comparable open source software suite for calculating NSAF values; i.e., TPP (Deutsch et al., 2010) combined with ABACUS (Fermin et al., 2011), using the proteasome spike-in experiment files as input. The .raw files were converted to

.mzML files by TPP Build 201411201551-6764 and then searched using the multi-threaded X!Tandem MS/MS search engine (Craig and Beavis, 2004) with the search parameters adjusted as close as possible to that used for the Morpheus searches. A decoy database was generated using the TPP tool DecoyFASTA for use with X!Tandem. Relevant X!Tandem parameters are listed here: parent monoisotopic mass error =  $\pm$  2.1 Da, fragment mass error =  $\pm$  0.025 Da, fixed modifications of carbamidomethylation (57.021464) on cysteine, and variable modification of oxidation (15.994915) on methionine, fully tryptic cleavages, missed cleavage sites = 2 maximum, no refinement and 8 threads. The configuration file used and the test datasets can be found in the Supporting Information. The data were analyzed in the TPP using Peptide Prophet (Keller et al., 2002). Relevant settings are listed: minimum probability = 0.05; minimum peptide length = 7; accurate mass, and nonparametric decoy database to pin down false discovery rate; ignore +1 charged spectra; and run Protein Prophet (Nesvizhskii et al., 2003) after Peptide Prophet. Once completed, all pepxml data from Peptide Prophet contained in a single folder was combined using the command line version of Protein Prophet from the TPP binaries with the following command: ProteinProphet.exe \*.pep.xml interact-COMBINED.prot.xml. This post-analysis aggregation was required for running the spectral counting program ABACUS. Here, we note that there are no graphic user interfaces to perform this

post analysis aggregation, which makes this portion of the data analysis more difficult for those unfamiliar with setting up and running programs from the command line. The combined data was analyzed by ABACUS with the following parameters: best peptide probability = 0.99; minimum peptide probability = 0.99; experimental peptide probability = 0; and combined file probability = 0.99 to most accurately match a 1% FDR stringency settings in MSpC. dNSAF values were compared in Microsoft Excel using the CORREL function and squaring the result. Additional timing tests were performed with a subset of the calibration curve data (ratios 0.091, 0.167, 0.333, and 0.500 in triplicate corresponding to 12 .raw files) by increasing file input to 24 (2x), 48 (4x), and 96 (8x) .raw files to determine the time dependence on input size between both pipelines tested (Morpheus/MSpC versus TPP/ABACUS). The timing tests and all data analyses were performed on a computer running Windows 7 Ultimate, with 16 GB of random access memory, and an Intel Core i7-2700k with hyper-threading turned on for eight logical cores.

#### **2.4. Discussion on Requiring Two Unique Peptides to Quantify a Protein**

Occasionally, some researchers may want to use a more stringent criterion for quantification such as requiring a protein to have more than one unique identifying peptide. To see how this might affect our data analysis, we re-analyzed our results

shown in Figure 2.3, this time requiring two unique peptides to quantify an individual protein. The results point to a very small increase in linearity observed in the average plots; however, there is a slight decrease in linearity for the egg sample (0.886 to 0.865) and a larger decrease in linearity for the individual UPS2 protein plot for the embryo sample (0.827 to 0.723). The decrease in linearity in the embryo sample is due to the analysis removing a low abundance UPS2 protein (O00762ups) identified with only one unique peptide. While some have suggested that requiring more than one unique peptide to identify a protein is an ideal approach, requiring two peptides for identifications in database searches reduces the number of protein identifications in the target database more than those in the decoy database and results in increased false discovery rates . While we recognize that researchers may want to implement more stringent requirements than what is typically used in database searching to quantify a set of proteins, there are two cases where requiring two unique peptides may not be ideal in a quantitative analysis. Firstly, low abundance proteins that have few PSMs might be identified by only a single peptide and thus be erroneously thrown out of the analysis. Secondly, there may be only one unique peptide that can differentiate between families of homologous proteins. In this second case, requiring two unique peptides would remove these homologous proteins from the MSpC analysis, even if they had a large number of

PSMs. Because of these reasons and because of the decreased linearity observed when requiring proteins to have two unique peptides (Figure 2.3) as compared to one unique peptide (Figure 2.2), we suggest caution in requiring more than one unique peptide per protein.

## **2.5. Tutorial**

# Morpheus Spectral Counter (MSpC) Tutorial

---

## Requirements

- 64 bit Windows Installation: For Example Microsoft Windows 7 64 bit, Windows 8 and 8.1 64bit, or Windows 10 64bit
    - This is due to requirements from vendor libraries to process Thermo .raw files directly in this tutorial
  - Download and install proteowizard (<http://proteowizard.sourceforge.net/downloads.shtml>)
    - Installation of proteowizard correctly installs the latest Thermo Vendor .dll's to read .raw files directly
  - Download pre-compiled binaries of revision 151 of the Morpheus Mass Spectrometry Search Engine (Thermo Version)  
([https://github.com/dcgemperline/Morpheus\\_SpC/releases/download/v1.0/Morpheus.Binaries.revision.151.zip](https://github.com/dcgemperline/Morpheus_SpC/releases/download/v1.0/Morpheus.Binaries.revision.151.zip)), referred from here on out as just Morpheus
  - Download and install MSpC  
([https://github.com/dcgemperline/Morpheus\\_SpC/releases/download/v1.0/MSpC\\_v1.0.zip](https://github.com/dcgemperline/Morpheus_SpC/releases/download/v1.0/MSpC_v1.0.zip))
    - MSpC depends on the latest .NET Runtime that will be installed with MSpC installer if it is not already installed
  - Download the data files from the PRIDE proteomics data repository using the following ID – PXD003002 Data Files (<http://www.ebi.ac.uk/pride/archive/>)
- 

## Data Analysis Starting from .raw files

1. Once all the requisite software is installed and the Data Files are downloaded you are ready to begin.
2. Unzip **Morpheus Binaries revision 151.zip** into your desired location and start **Morpheus (Thermo).exe** from the Morpheus (Thermo) folder.
3. Once Morpheus starts, verify at the top you are using **Morpheus (Thermo) revision 151**.
4. Add the following .raw files to Morpheus from the Raw Files folder contained in the Data Files folder (multiple files can be selected at once).
  - 04\_28\_15\_A\_rep1\_c1.raw
  - 04\_28\_15\_A\_rep2\_c1.raw

- 04\_28\_15\_A\_rep3\_c1.raw
  - 04\_28\_15\_B\_rep1\_c1.raw
  - 04\_28\_15\_B\_rep2\_c1.raw
  - 04\_28\_15\_B\_rep3\_c1.raw
  - 04\_28\_15\_C\_rep1\_c1.raw
  - 04\_28\_15\_C\_rep2\_c1.raw
  - 04\_28\_15\_C\_rep3\_c1.raw
  - 04\_28\_15\_D\_rep1\_c1.raw
  - 04\_28\_15\_D\_rep2\_c1.raw
  - 04\_28\_15\_D\_rep3\_c1.raw
5. Side Note on Fractionation: If you would like Morpheus, and thus MSpC to output summaries for a set of data by adding up all of the spectra identified in that dataset(as the case may be for fractionation data), place the files you would like to be summarized in seperate folders, such as SampleSet1, Sampleset2, and Morphues will generate protein\_groups.tsv and PSMs.tsv specifically for each sampleset. Later in MSpC after selecting the summary.tsv file these will show up as SampleSet1\*, SampleSet2\*, etc.
  6. Browse for the fasta file contained in the Data Files Folder  
**uniprot\_k12\_e\_coli\_contams\_plus\_proteasome\_and\_interactors.fasta**.
  7. Verify that Create Target-Decoy Database On The Fly is checked.
  8. Browse for an Output Folder. In this tutorial select browse, leave the default Desktop location highlighted and press Make a New Folder and rename it Morpheus Analysis. Press OK. You should now have a folder on the desktop called Morpheus Analysis, and Morpheus should say you are outputting data to the following output folder, where **NAME** is your username on your machine.
    - C:\Users\NAME\Desktop\Morpheus Analysis
  9. Set the Maximum threads to 2, for a dual core processor, 4 for a quad core processor, 8 for a quad core processor with hyperthreading, and 8 for an 8 core processor. These are reasonable values that will give you decent performance with Morpheus.
  10. Press Search, and the progress bar will indicate search progress for each raw file. On an Intel Core i7 2700K with Morpheus set to use 8 threads, this takes approximately 4.5 minutes.
  11. Close Morpheus if so desired.
  12. On completion of the search open MSpC.
  13. The indicators in Summary Directory Contains Valid Input should be red.
  14. Press the Select the Summary File button and navigate to the summary.tsv file contained in the Morpheus Analysis folder on the Desktop.
    - C:\Users\NAME\Desktop\Morpheus Analysis\summary.tsv
  15. The indicators in Summary Directory Contains Valid Input should now read green indicating that the Morpheus Analysis folder contains all the necessary files from Morpheus to run. (The files required to be in the same directory as summary.tsv are protein\_groups.tsv and PSMs.tsv. This will automatically occur if you output all of your Morpheus output into a single output folder.)

16. (Optionally) Select the provided whitelist file contained in the Data Files folder that you downloaded to simplify the output.
  17. Select your desired PSM FDR and Protein FDR (although the default options of 1% are good default options).
  18. Check or uncheck boxes of the desired calculations (NSAF, dNSAF, uNSAF)
  19. (Optionally) Select an Output Directory, otherwise MSpC defaults to the current directory of summary.tsv as an output directory.
    - Let's press Select Output Directory then Make a New Folder here called MSpC Output, press OK.
  20. Press Summarize Data, and the progress bar will indicate the progress of MSpC. This should take ~ 30 seconds or less
  21. The MSpC Output folder will then contain NSAF summaries for each individual .raw file, as well as a summary for all .raw files analyzed. These files are in the tab delimited output format .tsv
- 

## **Data Analysis Starting from previously searched Morpheus Data**

1. It is strongly suggested, that you start with the Data Analysis from .raw files tutorial first, and then refer back to the tutorial for previously searched Morpheus data if necessary
2. Open MSpC
3. Select a Morpheus summary.tsv file in a directory that contains the entire output from Morpheus.
4. If all output is there, the indicators in Summary Directory Contains Valid Input should now read green indicating that the folder contains all the necessary files from Morpheus to run MSpC.
5. (Optionally) Select the provided whitelist file contained in the Data Files folder that you downloaded to simplify the output.
6. Select your desired PSM FDR and Protein FDR (although the default options of 1% are good default options).
7. Check or uncheck boxes of the desired calculations (NSAF, dNSAF, uNSAF)
8. (Optionally) Select an Output Directory, otherwise MSpC defaults to the current directory of summary.tsv as an output directory.
  - Let's press Select Output Directory then Make a New Folder here called MSpC Output, press OK.
9. Press Summarize Data, and the progress bar will indicate the progress of MSpC. This should take ~ 30 seconds or less
10. The MSpC Output folder will then contain NSAF summaries for each individual .raw file, as well as a summary for all .raw files analyzed. These files are in the tab delimited output format .tsv

## 2.6. Acknowledgements

D.C.G. was supported by a grant from the U.S. Department of Energy Office of Science; Office of Basic Energy Sciences; Chemical Sciences, Geosciences, and Biosciences Division (DE-FG02-88ER13968) and a graduate training fellowship from the NIH (5 T32 GM 7133-37). M.S. and L.M.S were supported by a grant from the National Institute of Health/National Institute of General Medical Sciences (1P50HG004942). The authors thank Erin Gemperline, Richard S. Marshall, and Josh Coon for critical reading of the manuscript, and additionally thank Derek Bailey for a critical code review.

## 2.7. References

- Book, A.J., Gladman, N.P., Lee, S.S., Scalf, M., Smith, L.M., and Vierstra, R.D.** (2010). Affinity purification of the arabidopsis 26s proteasome reveals a diverse array of plant proteolytic complexes. *J. Biol. Chem.* **285**(33):25554–25569.
- Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M.** (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlfq. *Mol. Cell. Proteomics* **13**(9):2513–2526.

- Craig, R. and Beavis, R.C.** (2004). Tandem: matching proteins with tandem mass spectra. *Bioinformatics* **20**(9):1466–1467.
- Deutsch, E.W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J.K., Martin, D.B., Nesvizhskii, A.I., and Aebersold, R.** (2010). A guided tour of the trans-proteomic pipeline. *Proteomics* **10**(6):1150–1159.
- Fermin, D., Basrur, V., Yocom, A.K., and Nesvizhskii, A.I.** (2011). Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* **11**(7):1340–1345.
- Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P.** (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem ms. *Proc. Natl. Acad. Sci. USA* **100**(12):6940–6945.
- Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R.** (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by ms/ms and database search. *Anal. Chem.* **74**(20):5383–5392.
- Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R.** (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**(17):4646–4658.

- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M.** (2002). Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1(5):376–386.
- Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D.J.** (2004). Multiplexed protein quantitation in *saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3(12):1154–1169.
- Russell, J.D., Scalf, M., Book, A.J., Ladror, D.T., Vierstra, R.D., Smith, L.M., and Coon, J.J.** (2013). Characterization and quantification of intact 26s proteasome proteins by real-time measurement of intrinsic fluorescence prior to top-down mass spectrometry. *PLoS One* 8(3):e58157.
- Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C.** (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms. *Anal. Chem.* 75(8):1895–1904.
- Tu, C., Li, J., Sheng, Q., Zhang, M., and Qu, J.** (2014). Systematic assessment

of survey scan and ms<sub>2</sub>-based abundance strategies for label-free quantitative proteomics using high-resolution ms data. *J. Proteome Res.* **13**(4):2069–2079.

**Vizcaino, J.A., Cote, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Perez-Riverol, Y., Reisinger, F., Rios, D., Wang, R., and Hermjakob, H.** (2013). The proteomics identifications (pride) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**(Database issue):D1063–9.

Vizcaino, Juan Antonio Cote, Richard G Csordas, Attila Dianes, Jose A Fabregat, Antonio Foster, Joseph M Griss, Johannes Alpi, Emanuele Birim, Melih Contell, Javier O'Kelly, Gavin Schoenegger, Andreas Ovelleiro, David Perez-Riverol, Yasset Reisinger, Florian Rios, Daniel Wang, Rui Hermjakob, Henning eng BB/I024204/1/Biotechnology and Biological Sciences Research Council/United Kingdom WT085949MA/Wellcome Trust/United Kingdom Research Support, Non-U.S. Gov't England 2012/12/04 06:00 Nucleic Acids Res. 2013 Jan;41(Database issue):D1063-9. doi: 10.1093/nar/gks1262. Epub 2012 Nov 29.

**Wenger, C.D. and Coon, J.J.** (2013). A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *J. Proteome Res.* **12**(3):1377–1386.

**Wong, J.W. and Cagney, G.** (2010). An overview of label-free quantitation methods in proteomics by mass spectrometry. *Methods Mol. Biol.* **604**:273–283.

**Yang, P., Fu, H., Walker, J., Papa, C.M., Smalle, J., Ju, Y.M., and Vierstra, R.D.** (2004). Purification of the arabidopsis 26s proteasome: biochemical and molecular analyses revealed the presence of multiple isoforms. *J. Biol. Chem.* **279**(8):6401–6413.

**Zhang, B., VerBerkmoes, N.C., Langston, M.A., Uberbacher, E., Hettich, R.L., and Samatova, N.F.** (2006). Detecting differential and correlated protein expression in label-free shotgun proteomics. *J. Proteome Res.* **5**(11):2909–2918.

**Zhang, Y., Wen, Z., Washburn, M.P., and Florens, L.** (2010). Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* **82**(6):2272–2281.

**Zybailov, B., Mosley, A.L., Sardiu, M.E., Coleman, M.K., Florens, L., and Washburn, M.P.** (2006). Statistical analysis of membrane proteome expression changes in *saccharomyces cerevisiae*. *J. Proteome Res.* **5**(9):2339–2347.

## COLOPHON

This document was typeset with L<sup>A</sup>T<sub>E</sub>X . It is based on the University of Wisconsin dissertation template created by William C. Benton (available at <https://github.com/willb/wi-thesis-template>).